A Sharp analog of Young's Inequality on S^N and Related Entropy Inequalities

E. A. Carlen[1] E. H. Lieb[2] M. Loss[1]

1. School of Mathematics, Georgia Tech, Atlanta GA 30332

2. Departments of Mathematics and Physics, Jadwin Hall, Princeton University, P.O. Box 708, Princeton NJ 08544

Abstract We prove a sharp analog of Young's inequality on S^N, and deduce from it certain sharp entropy inequalities. The proof turns on constructing a nonlinear heat flow that drives trial functions to optimizers in a monotonic manner. This strategy also works for the generalization of Young's inequality on R^N to more than three functions, and leads to significant new information about the optimizers and the constants.

Math reviews Classification Numbers: 43A15, 52A40, 82C40

Key words: Inqualities, entropy, optimizers, best constants

[1] Work partially supported by U.S. National Science Foundation grant DMS 03-00349.

[2] Work partially supported by U.S. National Science Foundation grant PHY-0139984.

1. Introduction

This paper concerns further generalizations of the generalized Young's inequality due to Brascamp and Lieb [5], which we now recall.

For any $N \geq M$, let $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N$ be any set of N non zero vectors spanning $\mathbb{R}^M$. Let $f_1, f_2, \ldots, f_N$ be any set on N non negative measurable functions on $\mathbb{R}$. Given numbers p_j with $1 \leq p_j \leq \infty$ for $j = 1, 2 \ldots, N$, form the vector

$$\vec{p} = (1/p_1, 1/p_2, \ldots, 1/p_N) \ , \tag{1.1}$$

and define

$$D(\vec{p}) = \sup \left\{ \frac{\int_{\mathbb{R}^M} \prod_{j=1}^N f_j(\vec{a}_j \cdot x) \mathrm{d}^N x}{\prod_{j=1}^N \|f_j\|_{p_j}} \ : \ f_j \in L^{p_j}(\mathbb{R}) \qquad j = 1, 2, \ldots, N \right\} \ . \tag{1.2}$$

A theorem in [5] reduces the computation of $D(\vec{p})$ to a *finite dimensional* variational problem: Let $\mathcal{G}$ denote the set of all centered Gaussian functions on $\mathbb{R}$. That is, $g(x) \in \mathcal{G}$ if and only if $g(x) = ce^{-(sx)^2/2}$ for some $s > 0$ and some constant c. Define $D_{\mathcal{G}}(\vec{p})$ by

$$D_{\mathcal{G}}(\vec{p}) = \sup \left\{ \frac{\int_{\mathbb{R}^M} \prod_{j=1}^N g_j(\vec{a}_j \cdot x) \mathrm{d}^N x}{\prod_{j=1}^N \|g_j\|_{p_j}} \ : \ g_j \in \mathcal{G} \qquad j = 1, 2, \ldots, N \right\} \ . \tag{1.3}$$

It is proved in [5] that $D(\vec{p}) = D_{\mathcal{G}}(\vec{p})$, and hence

$$\int_{\mathbb{R}^M} \prod_{j=1}^N f_j(\vec{a}_j \cdot x) \mathrm{d}^N x \leq D_{\mathcal{G}}(\vec{p}) \prod_{j=1}^N \|f_j\|_{p_j} \ , \tag{1.4}$$

This can be used to explicit compute sharp constants in certain cases. For instance, when $M = 2$ and $N = 3$, $D_{\mathcal{G}}(\vec{p})$ may be evaluated, and this gives the sharp constant in the classical Young's inequality for convolutions.*

The first part of this paper concerns a version of this generalized Young's inequality for functions on the sphere S^{N-1}. Our generalization was motivated by statistical mechanical considerations, and was devised to prove a sharp entropy inequality for probability denisities on S^{N-1} which is also presented below. There are by now several alternative proofs of the Brascamp–Lieb inequality for functions on $\mathbb{R}^N$ (e.g., [8], [2] and [3]), but none of these seem to be readily adaptable to the consideration of functions on S^{N-1}, and it was necessary to develop a new approach.

* The sharp constant in Young's inequality for convolutions was obtained by Beckner at the same time that Brascamp and Lieb obtained their more general result. Beckner's results do not address the case of more than three functions, which is the focus here.

The new approach, it turns out, leads to a very simple proof of the original theorem in $\mathbb{R}^N$, and enables us to strengthen the original theorem in several respects, clearing up some questions left open by the authors cited above. In particular, we resolve a conjecture of Barthe whose incisive work in [3] settled many of the questions about non negative optimizers for the variational problem (1.2). We also obtain new information on the constants. For any given choice of $\{\vec{a}_1, \ldots, \vec{a}_N\}$, we give an explicit formula for the (unique) choice of $\vec{p}$ that minimizes $D(\vec{p})$, as well as the minimum value, which we refer to as the "best best constant" in the generalized Young's inequality.

We shall proceed to these results along the path which led to them, and begin by recalling some facts that motivated the investigation of a spherical analog of (1.4).

Let μ_N denote the uniform Borel probability measure on $S^{N-1}(\sqrt{N})$, the sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$, and let Let γ_N denote the Gaussian probability measure

$$\mathrm{d}\gamma_N = (2\pi)^{-N/2} e^{-|\vec{v}|^2/2} \mathrm{d}^N v \ . \tag{1.5}$$

We can consider μ_N as a probability measure on $\mathbb{R}^N$, supported on $S^{N-1}(\sqrt{N})$, and then it is a familiar fact that for large values of N, $\mathrm{d}\gamma_N \approx \mathrm{d}\mu_N$. In the considerations that motivated our investigation, a vector

$$\vec{v} = (v_1, v_2, \ldots, v_N) \tag{1.6}$$

in $\mathbb{R}^N$ represents the velocities of N one dimensional particles. Under any sort of evolution of the particle system that conserves kinetic energy, $\sum_{j=1}^{N} v_j^2$ will be constant. Supposing that its initial value is N, at any later time the state of the system will be given by a point in $S^{N-1}(\sqrt{N})$. The uniform probability measure μ_N is called the *microcanonical ensemble* in statistical mechanics. The proability measure $\mathrm{d}\gamma_N$ on the other hand would be called the *canonical ensemble* for this system. The principle of *equivalence of ensembles* is a cornerstone of equilibrium statistical mechanics. For this simple system, it reduces to the statement that for any fixed positive integer k, and any bounded measurable function $\phi(v_1, v_2, \ldots, v_k)$ of the first k velocities only,

$$\lim_{N\to\infty} \left(\int_{S^{N-1}(\sqrt{N})} \phi(v_1, v_2, \ldots, v_k) \mathrm{d}\mu_N - \int_{\mathbb{R}^N} \phi(v_1, v_2, \ldots, v_k) \mathrm{d}\gamma_N \right) = 0 \ .$$

However, the equivalence of ensembles goes only so far. A fundamental qualitative difference between γ_N and μ_N is that under the first probability measure, the coordinate functions are independent random variables, while under the second they are not. This lack of independence has an important quantitative effect that *does not diminish with increasing* N if we consider functions of all of the velocities at once, as we now explain.

Before going further, it will be convenient to make a change of scale, and consider the unit sphere. The factors of $\sqrt{N}$ that are necessary for comparison to the Gaussian measure $\mathrm{d}\mu_N$ will not be helpful in the next paragraphs. Therefore, let μ denote the

uniform Borel probability measure on S^{N-1}, the unit sphere in $\mathbb{R}^N$. Let π_j be the jth coordinate function. That is,

$$\pi_j(\vec{v}) = v_j \in [-1,1]. \tag{1.7}$$

Consider functions f_j defined on the interval $[-1,1]$ and pull them back to the sphere via the coordinate function π_j, i.e., $f_j(\pi_j(\vec{v}))$. We denote this function also by f_j. It will be clear from the context which of these functions is meant.

Because $\sum_{j=1}^N \pi_j(\vec{v})^2 = 1$, the coordinate functions are not independent random variables, and hence, given N functions f_j on $[-1,1]$, the quantities

$$\int_{S^{N-1}} \left(\prod_{j=1}^N f_j \right) \mathrm{d}\mu \qquad \text{and} \qquad \prod_{j=1}^N \left(\int_{S^{N-1}} f_j \mathrm{d}\mu \right) \ . \tag{1.8}$$

need not be equal. In fact, simple examples show that it is possible for the integral on the left in (1.8) to diverge while all of the integrals on the right are finite. However, according to the following theorem, such a divergence is not possible if each f_j is square integrable. Indeed, the product of the L^2 norms of the f_j controls the integral of $\prod_{j=1}^N f_j$ in the strongest way that one could hope. In what follows, $\|\cdot\|_{L^p(S^{N-1})}$ will denote an L^p norm with respect to μ on S^{N-1}.

Theorem 1 *For all $N \geq 2$, given non–negative measurable functions f_j, $j = 1, 2, \ldots, N$, on $[-1,1]$,*

$$\int_{S^{N-1}} \left(\prod_{j=1}^N f_j(v_j) \right) \mathrm{d}\mu \leq \prod_{j=1}^N \|f_j\|_{L^p(S^{N-1})} \ . \tag{1.9}$$

for all $p \geq 2$. Moreover, for each $p < 2$, there exist functions f_j so that $\|f_j\|_{L^p(S^{N-1})} < \infty$ for each j, while the integral on the left side of (1.9) diverges. Finally, for every $p \geq 2$ and $N \geq 3$, there is equality in (1.9) if and only if each f_j is constant.

It is natural to refer to (1.9) as a spherical version of the generalized Young's inequality (1.4). The resemblance is accentuated if we write $v_j = \vec{e}_j \cdot \vec{v}$ where the $\vec{e}_j$ are the standard basis vectors in $\mathbb{R}^N$. The proof that we give for Theorem 1 can be adapted to prove a generalization in which other vectors other than the $\vec{e}_j$ are considered, but this is not needed for the application that we now describe.

The inequality (1.9) implies a sharp entropy inequality for probability densities F on S^{N-1}. Indeed, let F be any probability density on S^{N-1}, and then, for each $j = 1, 2, \ldots, N$, let f_j denote the conditional expectation of F given v_j. This is a linear operation, and we define the operator P_j by

$$P_j F = f_j \ .$$

In more analytic terms, f_j is the function on $[-1,1]$ so that for all bounded measurable functions ϕ on $[-1,1]$,

$$\int_{S^{N-1}} \phi(v_j) F(\vec{v}) \mathrm{d}\mu = \int_{S^{N-1}} \phi(v_j) f_j(v_j) \mathrm{d}\mu \ .$$

As is evident from the definition, for square integrable F, $f_j = P_j F$ is just the orthogonal projection in $L^2(S^{N-1})$ of F onto the subspace consisting of functions depending only on v_j; i.e., measurable with respect to the sigma algebra generated by π_j.

There is yet another relation worth bearing in mind. To explain, introduce the one dimensional marginal ν_N of μ: For $N \geq 3$, and any function ϕ on $[-1,1]$,

$$\int_{S^{N-1}} \phi(v_1) \mathrm{d}\mu = \int_{[-1,1]} \phi(v) \mathrm{d}\nu_N$$

where

$$\mathrm{d}\nu_N = \frac{|S^{N-2}|}{|S^{N-3}|}(1-v^2)^{(N-3)/2}\mathrm{d}v \ . \tag{1.10}$$

Here, $|S^{m-1}|$ denotes the surface area of the $m-1$ dimensional unit sphere in $\mathbb{R}^m$; and $|S^0| = 2$. Then, $f_j(v)\mathrm{d}\nu_N$ is the marginal distribution of v_j under $F(\vec{v})\mathrm{d}\mu$. Whenever we refer to the jth marginal f_j of a probability density F on S^{N-1}, we shall mean that f_j is related to F in exactly this manner.

Clearly, each of the f_j is a probability density on S^{N-1}. For any probability density F on S^{N-1} the entropy of F is $S(F)$ defined by

$$S(F) = \int_{S^{N-1}} F \ln F \mathrm{d}\mu \ , \tag{1.11}$$

and likewise, the entropy of the marginal is given by

$$S(f_j) = \int_{[-1,1]} f_j \ln f_j \mathrm{d}\nu_N = \int_{S^{N-1}} f_j \ln f_j \mathrm{d}\mu \ .$$

How do the entropies of the marginals f_j compare with the entropy of their parent density F? The following theorem provides an answer:

Theorem 2 *For all $N \geq 2$, given any probability density F on S^{N-1}, let f_j be the jth marginal of F for $j = 1, 2, \ldots, N$. Then*

$$\sum_{j=1}^{N} S(f_j) \leq 2S(F) \ , \tag{1.12}$$

and the constant 2 on the right side of (1.12) is the best possible.

The inequality (1.12) may be compared to the familiar *subadditivity of the entropy* inequality on $\mathbb{R}^N$: Let G be any probability density on $\mathbb{R}^N$ with respect to $\mathrm{d}\gamma_N$, and let g_j denote its jth marginal, which is obtained by integrating out all of the variables except v_j. In this case, there is no relation among the coordinate functions. Hence

$$\int_{\mathbb{R}^N} \prod_{j=1}^{N} g_j \mathrm{d}\gamma_N = \prod_{j=1}^{N} \left(\int_{\mathbb{R}^N} g_j \mathrm{d}\gamma_N \right) = 1 \ ,$$

so that $H = \prod_{j=1}^N g_j$ is another probability density on $\mathbb{R}^N$. Then by Jensen's inequality,

$$\begin{aligned} 0 &\leq \int_{\mathbb{R}^N} \left(\frac{G}{H}\right) \ln\left(\frac{G}{H}\right) H \mathrm{d}\gamma_N \\ &= \int_{\mathbb{R}^N} G \ln G \mathrm{d}\gamma_N - \int_{\mathbb{R}^N} G \ln H \mathrm{d}\gamma_N \\ &= \int_{\mathbb{R}^N} G \ln G \mathrm{d}\gamma_N - \sum_{j=1}^N \int_{\mathbb{R}^N} G \ln g_j \mathrm{d}\gamma_N \ , \end{aligned}$$

and there is equality if and only if $G = H$. Defining the entropy $S(G)$ of a density G realtive to $\mathrm{d}\gamma_N$ by $S(G) = \int_{\mathbb{R}^N} G \ln G \mathrm{d}\gamma_N$, this says

$$\sum_{j=1}^N S(g_j) \leq S(G) \tag{1.13}$$

with equality if and only if $G = H$. Note the difference between (1.13) and (1.12): The latter requires an extra factor of 2 on the right, independent of N. This is due to the dependence of the coordinate functions v_j resulting from the constraint $\sum_{j=1}^N v_j^2 = 1$.

The difference between (1.13) and (1.12) is especially striking given the close relation between $\mathrm{d}\mu$ and $\mathrm{d}\gamma_N$. The inequality in Theorem 2 does not depend on the radius of the sphere, since the uniform measure is normalized, and so Theorem 2 says that there is a *dimension independent* departure from the equivalence of ensembles as measured by subadditivity of the entropy.

This dimension independence would not be guessed by linearizing the inequality in Theorem 2 about $F = 1$; *it is a non–perturbative effect*. The natural perturbative calculation would suggest that the difference between (1.13) and (1.12) "washes out" with increasing N, as we now explain.

Consider a probability denisty F on S^{N-1} of the form

$$F = 1 + \varepsilon H$$

where H is bounded and orthogonal to 1 in $L^2(S^{N-1})$. Then $f_j = P_j F = 1 + \varepsilon h_j$ where $h_j = P_j H$. Of course h_j is also orthogonal to 1 in $L^2(S^{N-1})$.

A simple and frequently encountered computation gives us

$$S(F) = \frac{\varepsilon^2}{2} \|H\|^2_{L^2(S^{N-1})} + \mathcal{O}(\varepsilon^3) \qquad \text{and} \qquad \sum_{j=1}^N S(f_j) = \frac{\varepsilon^2}{2} \|h_j\|^2_{L^2(S^{N-1})} + \mathcal{O}(\varepsilon^3) \ .$$

Since

$$\|h_j\|^2_{L^2(S^{N-1})} = \langle P_j H, P_j H \rangle_{L^2(S^{N-1})} = \langle H, P_j H \rangle_{L^2(S^{N-1})} \ ,$$

if we define the operator

$$P = \frac{1}{N}\sum_{j=1}^{N} P_j \ ,$$

we have that

$$\frac{\sum_{j=1}^{N} S(P_j F)}{S(F)} = N\frac{\langle H, PH\rangle_{L^2(S^{N-1})}}{\|H\|^2_{L^2(S^{N-1})}} + \mathcal{O}(\varepsilon) \ .$$

An optimist might then hope that the supremum of $\sum_{j=1}^{N} S(P_j F)/S(F)$ taken over all probability densities F would be given by C_N where

$$C_N = \sup\left\{ N\frac{\langle H, PH\rangle_{L^2(S^{N-1})}}{\|H\|^2_{L^2(S^{N-1})}} \ : \ H \in L^2(S^{N-1}) \ , \ \langle H, 1\rangle_{L^2(S^{N-1})} = 0 \right\} \ . \qquad (1.14)$$

The computation of the supremum is an eigenvalue problem, and has been done in [6; Theorems 1.2 and 2.1]. The result is

$$C_N = 1 + \frac{3}{N+1} \ .$$

The surplus over 1, namely $3/(N+1)$, measures the "departure from independence" as a function of N. Thus, if one considers densities F that deviate only slightly from the uniform density, one gets a correction term to the constant 1 in the Gaussian entropy inequality (1.13) that "remembers" the dependence of the coordinates on the sphere, but which vanishes as $N \to \infty$.

The precise size of this "departure from independence" as a function of N is crucial in some problems of non–equilbrium statisitical mechanics. The computation of C_N was at the heart of recent progess in computing the rate of relaxation to equilibrium in kinetic theory by direct consideration of an N body system, as proposed long ago by Mark Kac. For more details, see [6] and [7].

The fact that for more general densities F, the correction term to the Gaussian entropy inequality (1.13) does not vanish as $N \to \infty$ complicates the estimation of rates of realxation in entropic terms for N body systems in kinetic theory. This said, we turn to the proof of Theorem 2.

Proof of Theorem 2: Let F be any probability density on S^{N-1}, $N \geq 2$, and let f_j, $j = 1, 2, \ldots, N$ be its marginals. Then since f_j is a probability density, $\|f_j^{1/2}\|_{L^2(S^{N-1})} = 1$. As a consequence of Theorem 1, if we define C by

$$C = \int_{S^{N-1}} \left(\prod_{j=1}^{N} f_j^{1/2} \right) \mathrm{d}\mu \ ,$$

we have $C \leq 1$.

Suppose that $C = 0$. Then, $\prod_{j=1}^{N} f_j = 0$ almost everywhere, and so $\sum_{j=1}^{N} \ln f_j = -\infty$ almost everywhere. This would imply

$$-\infty = \int_{S^{N-1}} F\left(\sum_{j=1}^{N} \ln f_j\right) \mathrm{d}\mu = \sum_{j=1}^{N} S(f_j) \ .$$

This is impossible, since by Jensen's inequality, $S(f_j) \geq 0$ for each j. Therefore, $0 < C < 1$, and we may define a probability density H on S^{N-1} through

$$H = \frac{1}{C} \prod_{j=1}^{N} f_j^{1/2} \ .$$

As above, we now apply Jensen's inequality to conclude that

$$0 \leq \int_{S^{N-1}} \left(\frac{F}{H}\right) \ln\left(\frac{F}{H}\right) H \mathrm{d}\mu \ .$$

The right and side is easily seen to be

$$\int_{S^{N-1}} F \ln F \mathrm{d}\mu - \int_{S^{N-1}} F \ln H \mathrm{d}\mu =$$

$$\int_{S^{N-1}} F \ln F \mathrm{d}\mu - \sum_{j=1}^{N} \int_{S^{N-1}} F \ln f_j^{1/2} \mathrm{d}\mu + \ln(C) =$$

$$\int_{S^{N-1}} F \ln F \mathrm{d}\mu - \frac{1}{2} \sum_{j=1}^{N} \int_{S^{N-1}} f_j \ln f_j \mathrm{d}\mu + \ln(C) \ .$$

Since $\ln(C) < 1$ unless each $f_j = 1$, the inequality is proved, with equality holding only when each $f_j = 1$. The fact that the constant cannot be less than 2 in the inequality follows by finding a trial function that we present in the Appendix. ■

As discussed above, the factor of 2 in Theorem 2 is a correction to the classical subadditivity of the entropy that is required on account of the dependence of the coordinate functions due to the constraint $\sum_{j=1}^{N} v_j^2 = 1$. The remarkable fact that the size of this effect is independent of N depends on the specific nature of the constraint, and is not a general fact.

For example, consider the planar constraint $\sum_{j=1}^{N} v_j = 0$ on $\mathbb{R}^N$, and let P_{N-1} denote the hyperplane specified by this constraint. Let $\tilde{\mu}$ be a centered, isotropic Gaussian

probability measure on P_{N-1}. As we explain below, a special case of the Brascamp–Lieb Theorem yields the sharp inequalitiy

$$\int_{P_{N-1}} \left(\prod_{j=1}^{N} f(v_j) \right) \tilde{\mu} \leq \prod_{j=1}^{N} \|f_j\|_{L^{N/(N-1)}(P_{N-1})} \ . \tag{1.15}$$

This is an analog of (1.9) for the planar constraint. Notice however, that this time the L^p indices depend on N, and diminish towards 1 as N increases.

Just as Theorem 2 follows from Theorem 1, one obtains an entropy subadditivity inequality for the planar constraint from (1.15). Given a probability density F on P_{N-1} with respect to the reference measure $\tilde{\mu}$, define the marginal densities as above. Then the analog of (1.12) is the inequality

$$\sum_{j=1}^{N} S(f_j) \leq \frac{N}{N-1} S(F) \ . \tag{1.16}$$

This time, since $\lim_{N\to\infty} N/(N-1) = 1$, the effect of the contraint, as far as subadditivity of the entropy is concerned, diminishes to zero as N tends to infinity.

The connection between (1.15) and Young's inequality is revealing. To see the connection, we change of variables. Let $\vec{e}_j$, $j = 1, 2, \ldots, N$ be the standard basis vectors in $\mathbb{R}^N$. Let $\vec{u}_j$ be the normalized orthogonal projection of $\vec{e}_j$ onto the hyperplane P_{N-1}. One easily checks that for $i \neq j$,

$$\vec{u}_i \cdot \vec{u}_j = -\frac{1}{N-1}$$

and that for $\vec{v}$ in P_{N-1},

$$v_i = \vec{v} \cdot \vec{e}_i = \sqrt{\frac{N-1}{N}} \vec{v} \cdot \vec{u}_j \tag{1.17}$$

and

$$\sum_{j=1}^{N} |\vec{u} \cdot \vec{v}|^2 = \frac{N}{N-1} |\vec{v}|^2 \ . \tag{1.18}$$

For convenient constants, choose a scale so that the Gaussian denisty is $M(\vec{v}) = e^{-\pi |\vec{v}|^2}$ Defining the single variable funtions $g_j(y) = f_j \left(\sqrt{\frac{N-1}{N}} y \right) e^{-\pi (N-1) y^2/N}$, we have from (1.17) and (1.18) that

$$\prod_{j=1}^{N} f_j(v_j) M(\vec{v}) = \prod_{j=1}^{N} g_j(\vec{u}_j \cdot \vec{v}) \ , \tag{1.19}$$

and if $\mathrm{d}\mathcal{L}$ denotes Lebesgue measure on P_{N-1},

$$\int_{P_{N-1}} \left(\prod_{j=1}^{N} f_j(v_j) \right) \mathrm{d}\tilde{\mu} = \int_{P_{N-1}} \left(\prod_{j=1}^{N} g_j(\vec{u}_j \cdot \vec{v}) \right) \mathrm{d}\mathcal{L} \ . \tag{1.20}$$

Furthermore, for each j, $\int_{P_{N-1}} |f_j(v_j)|^{N/(N-1)} \mathrm{d}\tilde{\mu} = \int_{I\!R} |g_j(y)|^{N/(N-1)} \mathrm{d}y$. Therefore, identifying P_{N-1} with $I\!R^{N-1}$ in the obvious way, (1.15) is equivalent to the inequality

$$\int_{I\!R^{N-1}} \left(\prod_{j=1}^{N} g_j(\vec{u}_j \cdot \vec{v}) \right) \mathrm{d}\mathcal{L} \leq \prod_{j=1}^{N} \|g_j\|_{L^{N/(N-1)}(I\!R)} \ , \tag{1.21}$$

which is a special case of the Brascamp–Lieb generalization of Young's inequality.

In fact, for $N = 3$, it follows from the sharp form of the classical Young's inequality for convolutions. To see this, let $s = \vec{u}_1 \cdot \vec{v}$ and $t = \vec{u}_3 \cdot \vec{v}$, and notice that since $\vec{u}_1 + \vec{u}_2 + \vec{u}_3 = 0$, we have that $-(s+t) = \vec{u}_2 \cdot \vec{v}$. A simple computation reveals that $\mathrm{d}\mathcal{L} = (2/\sqrt{3})\mathrm{d}s\mathrm{d}t$, so that the $N = 3$ case of (1.21) becomes

$$\int_{I\!R^2} g_1(s) g_2(-s-t) g_3(t) \mathrm{d}s\mathrm{d}t \leq \frac{\sqrt{3}}{2} \|g_1\|_{L^{3/2}(I\!R)} \|g_2\|_{L^{3/2}(I\!R)} \|g_3\|_{L^{3/2}(I\!R)} \ .$$

This in turn is equivalent to the inequality

$$\|g_1 * g_2\|_{L^3(I\!R)} \leq \frac{\sqrt{3}}{2} \|g_1\|_{L^{3/2}(I\!R)} \|g_2\|_{L^{3/2}(I\!R)} \ ,$$

which is sharp. We now turn to the proof of Theorem 1.

2. Proof of Young's inequality on S^{N-1}

We prove Theorem 1 using a non-linear heat semigroup. For $1 \leq i, j \leq N$, $i \neq j$ let

$$L_{i,j} = v_i \frac{\partial}{\partial v_j} - v_j \frac{\partial}{\partial v_i} \ .$$

The Laplacian on S^{N-1} is the operator

$$\Delta = \sum_{i<j} L_{i,j}^2 = \frac{1}{2} \sum_{i \neq j} L_{i,j}^2 \ . \tag{2.1}$$

The normalization of the gradient on S^{N-1} implicit in this is convenient; for smooth functions f and g, we write

$$\nabla f \cdot \nabla g = \sum_{i<j} L_{i,j} f L_{i,j} g = \frac{1}{2} \sum_{i \neq j} L_{i,j} f L_{i,j} g \ , \tag{2.2}$$

and $|\nabla f|^2 = \nabla f \cdot \nabla f$.

Now fix any $p \geq 1$. For any smooth, non negative function g in $L^p(S^{N-1})$, and any $t > 0$, define

$$g(\vec{v}, t) = \left(e^{t\Delta} g^p(\vec{v})\right)^{1/p} \ . \tag{2.3}$$

The first thing to observe is that $g(\cdot, t)$ will be smooth and strictly positive for all $t > 0$, and the $L^p(S^{N-1})$ norm of g is conserved under this evolution:

$$\|g(\cdot, t)\|_{L^p(S^{N-1})} = \|g\|_{L^p(S^{N-1})} \tag{2.4}$$

for all $t \geq 0$.

The second thing to observe is that if g depends only on v_j for some j, so does $g(\cdot, t)$. The reason is that g depends only on v_j if and only if g is invariant under all rotations that fix the jth coordinate axis, and these rotations commute with the Laplacian. We write $g(v_j, t)$ to denote the evolution of such a function.

The third thing to observe is that the evolution, though non-linear, has the semigroup property: For all $s, t > 0$,

$$g(\vec{v}, s + t) = \left(e^{s\Delta} g^p(\vec{v}, t)\right)^{1/p} \ . \tag{2.5}$$

The fourth thing to observe is that

$$\lim_{t \to \infty} g(\vec{v}, t) = \|g\|_{L^p(S^{N-1})} \tag{2.6}$$

uniformly in $\vec{v}$.

Finally, a simple computation shows that for any smooth, positive function g on S^{N-1},

$$\left.\frac{\partial}{\partial t} g(v, t)\right|_{t=0} = \frac{1}{p} g^{1-p} \Delta g^p = \Delta g + (p-1)\frac{|\nabla g|^2}{g} \ . \tag{2.7}$$

Lemma 2.1 *Consider any N non negative functions $g_1, g_2, \ldots, g_N$ in $L^2([-1, 1], \mathrm{d}\nu_N)$. Use $p = 2$ and g_j in place of g in (2.3) to define $g_j(v_j, t)$. Then by the smooting properties of the heat equation, the function $\phi(t)$ defined by*

$$\phi(t) = \int_{S^{N-1}} \prod_{j=1}^{N} g_j(v_j, t) \mathrm{d}\mu$$

is differentiable for all $t > 0$, and is right continuous at $t = 0$. Moreover, introducing the functions h_k and G defined by

$$h_j(v_j, t) = \ln g_j(v_j, t) \qquad k = 1, 2 \ldots, N \qquad \text{and} \qquad G = \prod_{j=1}^{N} g_j \ , \tag{2.8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(t) = \frac{1}{2}\int_{S^{N-1}} \sum_{i\neq k} \left[(L_{i,k}h_k) - (L_{i,k}h_i)\right]^2 G\mathrm{d}\mu \ . \tag{2.9}$$

Proof The statements about smoothness and continuity require no justification. Taking $p = 2$ and $g = g_k(v_k, t)$ in (2.7), we have

$$\frac{\partial}{\partial t}g_k(v_k,t) = \Delta g_k(v_k,t) + \frac{|\nabla g_k(v_k,t)|^2}{g_k(v_k,t)} \ .$$

Hence, supressing the arguments on the right,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{S^{N-1}} \prod_{j=1}^{N} g_j(v_j,t)\mathrm{d}\mu\right)\Big|_{t=0} = \sum_{k=1}^{N}\int_{S^{N-1}} \left(\Delta g_k + \frac{|\nabla g_k|^2}{g_k}\right) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu \ .$$

The integral on the right can be written as

$$\int_{S^{N-1}} \sum_{k=1}^{N} (\Delta g_k) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu + \int_{S^{N-1}} \sum_{k=1}^{N} \left(\frac{|\nabla g_k|^2}{g_k}\right) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu \ . \tag{2.10}$$

clearly, the second integral on the right is non negative. We therefore examine the first integral.

Observe that $L_{i,j}g_k = 0$ unless either $i = k$ or $j = k$. Therefore,

$$\int_{S^{N-1}} \sum_{k=1}^{N} (\Delta g_k) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu = \int_{S^{N-1}} \sum_{k=1}^{N} \left(\sum_{i<k} L_{i,k}^2 g_k + \sum_{j>k} L_{k,j}^2 g_k\right) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu \tag{2.11}$$

Integrating by parts,

$$\int_{S^{N-1}} \sum_{k=1}^{N} \left(\sum_{i<k} L_{i,k}^2 g_k\right) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu = -\int_{S^{N-1}} \sum_{k=1}^{N} \left(\sum_{i<k} L_{i,k}g_k L_{i,k}g_i\right) \prod_{\ell=1,\ell\neq i,k}^{N} g_\ell \mathrm{d}\mu \ . \tag{2.12}$$

Using the notations in (2.8), the integral on the right side of (2.12) is

$$-\int_{S^{N-1}} \sum_{k=1}^{N}\sum_{i<k} \left(L_{i,k}h_k L_{i,k}h_i\right) G\mathrm{d}\mu \ .$$

Doing the same integration by parts on the remaining terms in (2.11), and substituting i for j, we have

$$\int_{S^{N-1}} \sum_{k=1}^{N} (\Delta g_k) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu = -\int_{S^{N-1}} \sum_{i\neq k} \left(L_{i,k}h_k L_{i,k}h_i\right) G\mathrm{d}\mu \ . \tag{2.13}$$

With the same notations, we have

$$\int_{S^{N-1}} \sum_{k=1}^{N} \left(\frac{|\nabla g_k|^2}{g_k} \right) \prod_{\ell=1,\ell\neq k}^{N} g_\ell \mathrm{d}\mu = \int_{S^{N-1}} \sum_{k=1}^{N} (\nabla h_k)^2 G \mathrm{d}\mu$$

$$= \int_{S^{N-1}} \sum_{i\neq k} (L_{i,k} h_k)^2 G \mathrm{d}\mu \tag{2.14}$$

$$= \frac{1}{2} \int_{S^{N-1}} \sum_{i\neq k} \left[(L_{i,k} h_k)^2 + (L_{i,k} h_i)^2 \right] G \mathrm{d}\mu \ .$$

Combining (2.13) and (2.14) we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\int_{S^{N-1}} \prod_{j=1}^{N} g_j(v_j, t) \mathrm{d}\mu \right) \Big|_{t=0} = \frac{1}{2} \int_{S^{N-1}} \sum_{i\neq k} \left[(L_{i,k} h_k) - (L_{i,k} h_i) \right]^2 G \mathrm{d}\mu \ .$$

This is (2.9). ■

Proof of Theorem 1: By Lemma 2.1, the difference between the right and left hand sides of (1.9) is

$$\int_0^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t} \int_{S^{N-1}} \prod_{j=1}^{N} g_j(v_j, t) \mathrm{d}\mu \right) \mathrm{d}t =$$

$$\frac{1}{2} \int_0^\infty \left(\int_{S^{N-1}} \sum_{i\neq k} \left[(L_{i,k} h_k(v_k, t)) - (L_{i,k} h_i(v_i, t)) \right]^2 G \mathrm{d}\mu \right) \mathrm{d}t \geq 0 \ .$$

This proves the inequality.

Also, it is now clear that for all $t > 0$, each h_k is smooth and bounded, and G is strictly positive, so that there is equality in (1.9) if and only if $\left[(L_{i,k} h_k) - (L_{i,k} h_i) \right]^2 = 0$ for all $t > 0$, all $\vec{v}$ and all $i \neq k$.

Fixing t, i and k, this requires

$$v_i h_k'(v_k) = -v_k h_i'(v_i) \tag{2.15}$$

This implies that for some constant c, $\begin{bmatrix} h_i'(v_i) \\ h_k'(v_k) \end{bmatrix} = c \begin{bmatrix} -v_i \\ v_k \end{bmatrix}$ for all values of v_i and v_k. Hence, for all $i \neq k$, h_i' and h_k' are linear functions with slopes of the same magnitude but opposite signs. For $N \geq 3$, the signs of all pairs cannot be opposite unless all of the slopes

are zero. This concludes the proof that there is equality in (1.9) if and only if each of the functions there is constant.*

In the appendix, there is an explicit example showing that (1.9) cannot hold if $L^2(S^{N-1})$ is replaced by $L^p(S^{N-1})$ for any $p < 2$. In fact, it is shown that for any $p < 2$, there is a function f so that with $f_j = f$ for all j, the left hand side of (1.9) is infinite, and the right hand side is finite. Alternatively, one can see that if (1.9) did hold with 2 repalced by some $p < 2$, then Theorem 2 would hold with 2 replaced by this value of p, which we have seen is not possible. ■

The simple heat flow argument that was used to prove Theorem 1 can be adapted to other situations as well. Indeed, one could easily consider inequalities for integrals over S^{N-1} of more general products $\prod_{j=1}^{P} f_j(\vec{a}_j \cdot \vec{v})$. The case considered here was $P = N$ and $\vec{a}_j = \vec{e}_j$ because that was what was relevant for Theorem 2. Further generalizations are possible, and may be interesting.

In the next sections, we exhibit the versatility of the method by showing that a heat flow interpolation between trial functions and Gaussians can be used to prove the original Brascamp Lieb inequality on $\mathbb{R}^M$. Barthe [2],[3] has given a proof of this inequality, together with a dual inverse inequality using an interpolation based on optimal mass transport. It was somewhat surprising that one could prove the Brascamp Lieb inequality with such a simple heat flow interpolation, and after hearing a report on our work, Barthe and Cordero–Erausquin [4] added to the surprise by showing that a heat flow interpolation could be used to prove the inverse dual inequality as well.

3. The generalized Young's inequality on $\mathbb{R}^M$

We have introduced this inequality in the introduction, and shall use the same notation here. Recall that for any $N \geq M$, $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N$ is a set of N non zero vectors in $\mathbb{R}^M$. Let $f_1, f_2, \ldots, f_N$ be any set of N non negative measurable functions on $\mathbb{R}$, and consider the integral

$$\int_{\mathbb{R}^M} \prod_{j=1}^{N} f_j(\vec{a}_j \cdot x) \mathrm{d}^M x \ . \tag{3.1}$$

There are certain natural restrictions on the underlying set of vectors $\vec{a}_j$. First of all, $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ must span $\mathbb{R}^M$ for the integral in (1.2) to possibly converge. Second, it is natural to assume that no pair of vectors $\vec{a}_i$ and $\vec{a}_j$ are proportional; if they were, we could combine two factors into one in the integrand (1.2). *These assumptions will be in force throughout the following sections.*

As before, given numbers p_j with $1 \leq p_j \leq \infty$ for $j = 1, 2 \ldots, N$, form the vector $\vec{p} = (1/p_1, 1/p_2, \ldots, 1/p_N)$, and define $D(\vec{p})$ and $D_{\mathcal{G}}(\vec{p})$ through (1.2) and (1.3) respectively.

* The analysis of (2.15) in a preprint of this paper contained an error. This was pointed out and corrected in a private communication from Shannon Starr, to whom we are grateful.

The Brascamp and Lieb Theorem asserts that $D(\vec{p}) = D_{\mathcal{G}}(\vec{p})$. As in the proof of Theorem 2, we shall use a non linear semigroup based on an appropriately chosen heat kernel to interpolate between arbitrary trial functions and Gaussian optimizers. The appropriate choice of the heat kernel depends on both $\vec{p}$ and the vectors $\{\vec{a}_1, \ldots, \vec{a}_N\}$. We shall show in this section that the desired heat kernel exist whenever the supremum is attained the Gaussian variational problem (1.3) for given $\vec{p}$ and $\{\vec{a}_1, \ldots, \vec{a}_N\}$. Note that the supremum being attained means that there are numbers $0 < s_j < \infty$ so that with $g_j(y) = e^{-(s_j^2 y^2)/2}$,

$$\int_{I\!R^M} \prod_{j=1}^N g_j(\vec{a}_j \cdot x) \mathrm{d}^N x = D_{\mathcal{G}}(\vec{p}) \prod_{j=1}^N \|g_j\|_{p_j}$$

In this case, we shall say that the Gaussian variational problem has an optimizer, and we identify the optimizer with the vector in $I\!R^N$ whose jth entry is s_j.

Theorem 3.1 *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be a set of vectors spanning $I\!R^M$ and suppose that the vector $\vec{p}$ is such that the Gaussian variational problem (1.3) has a maximizer. Then $D(\vec{p}) = D_{\mathcal{G}}(\vec{p})$.*

By itself, this theorem is contained in the Brascamp–Lieb Theorem, which asserts that $D(\vec{p}) = D_{\mathcal{G}}(\vec{p})$ in general. However, as we shall see in the next section, Theorem 3.1 provides the essential reduction of (1.2) to (1.3), and to complete the analysis and prove the full result, one needs only certain facts about the Gaussian variational problem. For the most part, the facts we need are contained in the work of Barthe [3], so that once we have proved Theorem 3.1, our work is largely done. The rest of this section is devoted to the proof of Theorem 3.1.

As preparation for the proof, let R be any invertible $M \times M$ matrix, and consider the heat semigroup e^{tL} generated by

$$L = \nabla \cdot R^t R \nabla \ . \tag{3.2}$$

For each j, and each $t > 0$, define $f_j(\cdot, t)$ by

$$f_j(t, \vec{a}_j \cdot x) = \left(e^{tL} f_j^{p_j}(\vec{a}_j \cdot x)\right)^{1/p_j} \ . \tag{3.3}$$

Since L commutes with translations, the set of functions on $I\!R^M$ of the form $f(\vec{a} \cdot x)$ is invariant under e^{tL} . In fact, for any bounded function f_0 on $I\!R$, for all $t > 0$, $e^{tL} f_0(\vec{a} \cdot x) = f(t, \vec{a} \cdot x)$ where $f(t, y)$ is the solution of

$$\frac{\partial}{\partial t} f(t, y) = |R\vec{a}|^2 \frac{\partial^2}{\partial y^2} f(t, y) \qquad f(0, y) = f_0(y) \ . \tag{3.4}$$

The fundamental solution of (3.4) is $g(t, y) = \dfrac{1}{\sqrt{4\pi |R\vec{a}|^2 t}} e^{-y^2/(4|R\vec{a}|^2 t)}$. Therefore, with pointwise convergence,

$$\lim_{t \to \infty} t^{1/2p_j} f_j(t, t^{1/2} y) = \frac{\|f_j\|_{p_j}}{(4\pi |R\vec{a}|^2)^{1/2p_j}} e^{-y^2/(4|R\vec{a}|^2 p_j)} \ . \tag{3.5}$$

Let $g_j(y)$ denote the centered Gaussian function defined by the right hand side of (3.5). Note also that for each j and t,

$$\|f_j(t,\cdot)\|_{p_j} = \|f_j\|_{p_j} = \|g_j\|_{p_j} \ . \tag{3.6}$$

If we assume that each f_j is bounded and has compact support, then it is possible to obtain simple Gaussian bounds on each $f_j(t,y)$ from which, using (3.5) and the obvious dominated convergence argument, it follows that

$$\lim_{t\to\infty} \int_{I\!R^M} \prod_{j=1}^N t^{1/2p_j} f_j(t, t^{1/2}(\vec{a}_j \cdot x)) \mathrm{d}^M x = \int_{I\!R^M} \prod_{j=1}^N g_j(\vec{a}_j \cdot x)) \mathrm{d}^M x \ .$$

Moreover, by the scale invariance that obtains under (4.1),

$$\int_{I\!R^M} \prod_{j=1}^N t^{1/2p_j} f_j(t, t^{1/2}(\vec{a}_j \cdot x)) \mathrm{d}^M x = \int_{I\!R^M} \prod_{j=1}^N f_j(t, (\vec{a}_j \cdot x)) \mathrm{d}^M x$$

so that

$$\lim_{t\to\infty} \int_{I\!R^M} \prod_{j=1}^N f_j(t, (\vec{a}_j \cdot x)) \mathrm{d}^M x = \int_{I\!R^M} \prod_{j=1}^N g_j(\vec{a}_j \cdot x)) \mathrm{d}^M x \ .$$

It now follows that *if* we can choose R so that $\displaystyle\int_{I\!R^M} \prod_{j=1}^N f_j(t, (\vec{a}_j \cdot x)) \mathrm{d}^M x$ is a non decreasing function of t, then

$$\frac{\int_{I\!R^M} \prod_{j=1}^N f_j(\vec{a} \cdot x) \mathrm{d}^N x}{\prod_{j=1}^N \|f_j\|_{p_j}} \leq \frac{\int_{I\!R^M} \prod_{j=1}^N g_j(\vec{a} \cdot x) \mathrm{d}^N x}{\prod_{j=1}^N \|g_j\|_{p_j}} \ .$$

By this argument, proof of the Brascamp–Lieb Theorem is reduced to finding a fixed matrix R so that the function $\eta(t)$ defined by

$$\eta(t) = \int_{I\!R^M} \prod_{j=1}^N f_j(t, \vec{a}_j \cdot x) \mathrm{d}^N x \tag{3.7}$$

is non–decreasing where $f_j(t,y)$ is determined through the choice of R by (3.2) and (3.3).

If this is to work at all, the Gaussian functions g_j defined by the limit in (3.5) must be maximizers for the variational problem (1.2), and certainly for the variational problem (1.3). We can gain insight into how R must be chosen by considering the Euler–Lagrange equation for (1.3).

For each j, let $\tilde{g}_j$ be the centered Gaussian function given by $\tilde{g}_j(y) = e^{-(s_j y_j)^2/2}$. Then a simple calculation reveals that

$$2\ln\left(\frac{\int_{I\!R^M}\prod_{j=1}^N \tilde{g}_j(\vec{a}_j\cdot x)\mathrm{d}^M x}{\prod_{j=1}^N \|\tilde{g}_j\|_{p_j}}\right) = \sum_{j=1}\frac{1}{p_j}2\ln(s_j) - \ln\left(\det\left(AS^2A^t\right)\right) \ , \tag{3.8}$$

where S is the diagonal $N\times N$ matrix whose jth diagonal entry is s_j, and A is the $M\times N$ matrix whose jth column is $\vec{a}_j$; i.e., $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N]$. (This notation $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N]$ will be used repeatedly in what follows.)

Introduce the variables $t_1, t_2, \ldots, t_N$ by $t_j = \ln(s_j^2)$. Let T be the diagonal matrix whose j diagonal entry is t_j.

Define the function ϕ on $I\!R^N$ by

$$\phi(t_1, t_2, \ldots, t_N) = Tr\left(\ln\left(Ae^T A^t\right)\right) \ . \tag{3.9}$$

Since $\ln\left(\det\left(Ae^T A^t\right)\right) = Tr\left(\ln\left(Ae^T A^t\right)\right)$, we have from (1.3) and (3.8) that

$$2\ln(D_{\mathcal{G}}(\vec{p})) = \sup_{t_1,t_2,\ldots,t_N}\left\{\sum_{j=1}\frac{1}{p_j}t_j - \phi(t_1,t_2,\ldots,t_N)\right\} \tag{3.10}$$

A simple calculation shows that

$$\begin{aligned}\frac{\partial}{\partial t_j}\phi(t_1,t_2,\ldots,t_N) &= e^{t_j}\vec{a}_j\cdot(Ae^TA^t)^{-1}\vec{a}_j \\ &= (s_j\vec{a}_j)\cdot(AS^2A^t)^{-1}(s_j\vec{a}_j) \ .\end{aligned} \tag{3.11}$$

Therefore, the Euler–Lagrange equation for the optimization problem in (3.10) is

$$\begin{aligned}\frac{1}{p_j} &= s_j\vec{a}_j\cdot(AS(AS)^t)^{-1}s_j\vec{a}_j \\ &= \vec{e}_j\cdot(AS)^t(AS(AS)^t)^{-1}(AS)\vec{e}_j \ ,\end{aligned} \tag{3.12}$$

where $\vec{e}_j$ is the jth standard basis vector in $I\!R^N$. Notice that since $\mathrm{rank}(A) = M$, the matrix $(AS)^t(AS(AS)^t)^{-1}(AS)$ is just the orthogonal projection onto the range of AS.

We now show that if the supremum in (3.10) is a attained, so that there is a positive diagonal matrix S satisfying (3.12), then we can choose

$$R = (AS(AS)^t)^{-1/2} \ , \tag{3.13}$$

and with this choice, the function $\eta(t)$ defined by (3.7) is non–decreasing. The key is the following lemma:

Lemma 3.2 *Let $f_1, f_2, \ldots, f_N$ be N bounded, non–negative measurable functions on $\mathbb{R}$ with compact support. Let R be any invertible $M \times M$ matrix, and consider the heat semigroup e^{tL} generated by $L = \nabla \cdot R^t R \nabla$. For each j, and each $t > 0$, define $f_j(\cdot, t)$ by (3.3) and define Define the function $\eta(t)$ by (3.7). Then with $h_j(y,t) = \ln f_j(t,y)$ and $F(x,t) = \prod_{j=1}^N f_j(t, \vec{a}_j \cdot x)$, $\eta(t)$ is differentiable for $t > 0$, and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta(t) = \int_{\mathbb{R}^M} \left(\sum_{i,j=1}^M h_i'(\vec{a}_i \cdot x, t) Q_{i,j} h_j'(\vec{a}_j \cdot x, t) \right) F(x,t) \mathrm{d}^M x \ , \tag{3.14}$$

where Q is the $M \times M$ matrix with

$$Q_{i,j} = \delta_{i,j} p_j |R\vec{a}_j|^2 - R\vec{a}_i \cdot R\vec{a}_j \ . \tag{3.15}$$

Proof: By (3.4) we have that

$$\frac{\partial}{\partial t} f_j(t, \vec{a}_j \cdot x) = L f_j(t, \vec{a}_j \cdot x) + (p_j - 1) \frac{|R\nabla f_j(t, \vec{a}_j \cdot x)|^2}{f_j(t, \vec{a}_j \cdot x)}$$

and hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta(t) = \sum_{j=1}^N \int_{\mathbb{R}^M} \left[L f_j(\vec{a}_j \cdot x) + (p_j - 1) \frac{|R\nabla f_j(\vec{a}_j \cdot x, t)|^2}{f_j(\vec{a}_j \cdot x, t)} \right] \prod_{i \neq j} f_i(\vec{a}_i \cdot x, t) \mathrm{d}^M \ .$$

Let $h_j = \ln f_j$, and let $F(x) = \prod_{j=1}^N f_f(\vec{a}_j \cdot x)$. Then, integrating by parts in the term containing L, and suppressing arguments,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\eta(t) &= \int_{\mathbb{R}^M} \left[\sum_{j=1}^N (p_j - 1) |R\vec{a}_j|^2 |h_j'|^2 - \sum_{i \neq j} (R\vec{a}_i \cdot R\vec{a}_j) h_i' h_j' \right] F(x) \mathrm{d}^M x \\ &= \int_{\mathbb{R}^M} \left[\sum_{j=1}^N p_j |R\vec{a}_j|^2 |h_j'|^2 - \sum_{i,j} (R\vec{a}_i \cdot R\vec{a}_j) h_i' h_j' \right] F(x) \mathrm{d}^M x \end{aligned} \tag{3.16}$$

Using the definition (3.15), we have (3.14). ■

Proof of Theorem 3.1: We apply Lemma 3.2, and must choose R so that Q is a positive matrix. By assumption, there is a maximizer for the Brascamp–Lieb variational problem (1.3), or equivalently (3.10), and hence there is a positive diagonal matrix S

such that the Euler–Lagrange equation (3.12) is satisfied for each j. In this case with $R = (AS(AS)^t)^{-1/2}$, (3.15) becomes

$$Q = S^{-1}(I - P)S^{-1} \tag{3.17}$$

where $P = (AS)^t(AS(AS)^t)^{-1}(AS)$ is the orthogonal projection onto the range of AS. This is certainly non–negative, and hence whenever the Gaussian variational problem (1.3) has an optimizer, $D(\vec{p}) = D_{\mathcal{G}}(\vec{p})$. ■

We close this section by expressing $D(\vec{p})$ in terms of A and S when the supremum in (3.10) is attained. In this case, the optimizing Gaussians g_j are given by the limit in (3.5). We may assume that $\|f_j\|_{p_j} = 1$ for each j. With R given by (3.13), the Euler–Lagrange equation (3.12) says $|R\vec{a}_j|^2 = 1/(s_j^2 p_j)$, and hence $g_j(\vec{a}_j \cdot x) = \left(\frac{p_j s_j^2}{4\pi}\right)^{\frac{1}{2p_j}} e^{-(s_j \vec{a}_j \cdot x)^2/4}$. Thus,

$$\prod_{j=1}^N g_j(\vec{a}_j \cdot x) = \prod_{j=1}^N \left(\frac{p_j s_j^2}{4\pi}\right)^{\frac{1}{2p_j}} e^{-|ASx|^2/4} \ , \tag{3.18}$$

and therefore,

$$\begin{aligned} D(\vec{p}) &= \prod_{j=1}^N \left(\left(p_j s_j^2\right)^{1/(2p_j)} \right) \left(\frac{1}{4\pi}\right)^{M/2} \int_{I\!R^M} e^{-|ASx|^2/4} \mathrm{d}^M x \\ &= \prod_{j=1}^N \left(\left(p_j s_j^2\right)^{1/(2p_j)} \right) \det(AS^2A^t)^{-1/2} \ . \end{aligned} \tag{3.19}$$

For future use, note that (3.18) can be written as

$$\prod_{j=1}^N g_j(\vec{a}_j \cdot x) = D(\vec{p}) \left(\frac{1}{\int_{I\!R^M} e^{-|ASx|^2/4} \mathrm{d}^M x} \right) e^{-|ASx|^2/4} \ . \tag{3.20}$$

Note that if S satisfies the Euler–Lagrange equation (3.12). so does λS for any $\lambda > 0$. Replacing S by λS in (3.20), and taking λ to infinity, we obtain $D(\vec{p})\delta_0$ in the limit, where δ_0 is the point mass at the origin. This will be used later on.

4. The Gaussian optimization problem

The analysis in the previous section leads very naturally to the following questions:

• *For which values of $\vec{p}$ is $D_{\mathcal{G}}(\vec{p})$ finite?*

• *For which values of $\vec{p}$ is there an optimizer for the Gaussian variational problem (1.3)?*

These questions have been answered by Barthe [3]. (In the special case in which every set of M vectors chosen from among $\{\vec{a}_1, \ldots, \vec{a}_N\}$ is a basis, this had been done in [5]). The answers may appear unfavorable for our program, since it turns out that in general there exist $\vec{p}$ for which $D_{\mathcal{G}}(\vec{p})$ is finite, but for which there is no optimizer for the Gaussian variational problem. Hence one additional observation is required to deduce the Brascamp Lieb Theorem from Theorem 3.1.

First, we recall Barthe's answer to the first question, which is pleasingly simple: Let K_A denote the convex hull of the vectors $\vec{z}$ whose entries are either zero or one, and for which the set $\{\vec{a}_j \ : \ z_j = 1 \ \}$ is a basis of $\mathbb{R}^M$. Barthe has proved [3] that $D_{\mathcal{G}}(\vec{p})$ finite if and only if $\vec{p} \in K_A$.

Note that K_A lies in the hyperplane in $\mathbb{R}^N$ given by the equation $\sum_{j=1}^N z_j = M$. Let K_A° denote the interior of K_A relative to this hyperplane. Barthe has also proved in [3] that when $\vec{p} \in K_A^\circ$, the supremum in the Gaussian optimization problem (1.3) is attained.

In this section, we give another proof of these results. We do this for two reasons. First, our proof gives an alternate characterization of K_A that is directly checkable. Second, our proof is based on a *partial scale invariance* property of the functional that we seek to optimize. This partial scale invariance property of the functional is expressed in the identity (4.3) below. As we shall see, it completely determines the nature of K_A, and it provides a crucial handle on the variational problem in case $\vec{p}$ is on the boundary of K_A.

The obvious scale invariance argument shows that for $D(\vec{p})$ or even $D_{\mathcal{G}}(\vec{p})$ to be finite, it is necessary that

$$\sum_{j=1}^N \frac{1}{p_j} = M \ , \tag{4.1}$$

and of course that $1 \leq p_j \leq \infty$ for each j. Indeed, let λ be any positive number, and replace each $f_j(y)$ in (1.2) by $f_j(\lambda y)$. The numerator in (1.2) is proportional to $\lambda^{-\sum_{j=1}^N 1/p_j}$, while the denominator is proportional to λ^{-M}. This excludes a finite maximum unless (4.1) holds.

A somewhat less obvious *partial scaling argument* leads to further restrictions on $\vec{p}$. This depends on a simple identity that is crucial in what follows:

Lemma 4.1 *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be a set of vectors spanning $\mathbb{R}^N$. Let S be any proper, non–empty subset of $\{1, 2, \ldots, N\}$, and let let $r = \dim(\mathrm{span}(\{\vec{a}_j \ : \ j \in S\}))$ Then there are explicitly computable sets of vectors $\{\vec{b}_j \ : \ j \in S\}$ and $\{\vec{c}_j \ : \ j \in S^c\}$ such that for any set of non negative functions f_j, each bounded and with compact support,*

$$\begin{aligned} &\int_{\mathbb{R}^M} \prod_{j=1}^N f_j(\vec{a}_j \cdot x) \mathrm{d}^M x = \\ &\int_{\mathbb{R}^r} \prod_{j \in S} f_j(\vec{b}_j \cdot y) \left(\int_{\mathbb{R}^{M-r}} \prod_{j \in S^c} f_j(\vec{b}_j \cdot y + \vec{c}_j \cdot z) \mathrm{d}^{M-r} z \right) \mathrm{d}^r y \ . \end{aligned} \tag{4.2}$$

Proof: Let $\{\vec{u}_1, \ldots, \vec{u}_r\}$ be an orthonormal basis for the span of $\{\vec{a}_j \ : \ j \in S\}$. Let $\{\vec{v}_{r+1}, \ldots, \vec{v}_M\}$ be an orthonormal basis for the orthogonal complement. Choose the sign of $\vec{v}_M$ so that $\det([\vec{u}_1, \ldots \vec{u}_r, \vec{v}_{r+1}, \ldots, \vec{v}_M]) = 1$. Let $U = [\vec{u}_1, \ldots \vec{u}_r]$, and let $V = [\vec{v}_{r+1}, \ldots, \vec{v}_N]$. Define $\vec{b}_j = U^t \vec{a}_j$ and $\vec{c}_j = V^t \vec{a}_j$. Likewise define $y = U^t x$ and $z = V^t x$. Then $\vec{a}_j \cdot x = \vec{b}_j \cdot \vec{y} + \vec{c}_j \cdot \vec{z}$, and for $j \in S$, $\vec{a}_j \cdot x = \vec{b}_j \cdot \vec{y}$. Since $\mathrm{d}^M x = \mathrm{d}^r y \mathrm{d}^{M-r} z$, and since, by construction, $\vec{c}_j = 0$ for $j \in S$, (4.2) follows immediately. ■

To apply this, we rescale in $\mathrm{span}(\{\vec{a}_j \ : \ j \in S\})$ alone: For each $j \in S$, replace f_j by $f_j^{(\lambda)}$ where $f_j^{(\lambda)}(y) = \lambda^{1/p_j} f_j(\lambda y)$. Then $\|f_j^{(\lambda)}\|_{p_j} = \|f_j\|_{p_j}$, so that this replacement does not affect the denominator in (1.2). Then:

$$\int_{I\!R^M} \prod_{j \in S} f_j^{(\lambda)}(\vec{a}_j \cdot x) \prod_{j \in S^c} f_j(\vec{a}_j \cdot x) \mathrm{d}^M x = \left[\lambda^{-r} \prod_{j \in S} \lambda^{1/p_j}\right] \int_{I\!R^r} \prod_{j \in S} f_j(\vec{b}_j \cdot y) \left(\int_{I\!R^{M-r}} \prod_{j \in S^c} f_j(\lambda^{-1} \vec{b}_j \cdot y + \vec{c}_j \cdot z) \mathrm{d}^{M-r} z \right) \mathrm{d}^r y \ . \tag{4.3}$$

We see that if

$$\sum_{j \in S} \frac{1}{p_j} > r = \dim(\mathrm{span}(\{\vec{a}_j \ : \ j \in S\})) \ ,$$

then the integral in (4.3) diverges as λ tends to $+\infty$. Since $\|f_j^{(\lambda)}\|_{p_j} = \|f_j\|_{p_j}$, this means that $D(\vec{p})$ is infinite in this case. These considerations justify the following definitions. *

Definition Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be a given set of vectors spanning $I\!R^M$. For each subset S of $\{1, 2 \ldots, N\}$, define

$$r(S) = \dim(\mathrm{span}(\{\vec{a}_j \ : \ j \in S\})) \ . \tag{4.4}$$

Let K_A denote the subset of $I\!R^N$ consisiting of vectors $\vec{z}$ such that $\sum_{j=1}^N z_j = 1$, $0 \leq z_j \leq 1$ for each j, and finally

$$\sum_{j \in S} z_j \leq r(S) \ . \tag{4.5}$$

Define K_A° to be the subset of K_A consisting of vectors $\vec{z}$ satisfying

$$\sum_{j \in S} z_j < r(S) \tag{4.6}$$

* Note that as λ tends to zero, $\prod_{j \in S^c} f_j(\lambda^{-1} \vec{b}_j \cdot y + \vec{c}_j \cdot z)$ tends to zero, and will even vanish idenitically for λ large enough when the f_j have compact support. Hence, (4.3) does not give us information on the relation between $\sum_{j \in S} 1/p_j$ and $r(S)$ in the limit as λ tends to zero.

for all proper, non–empty subsets S of $\{1, 2 \ldots, N\}$. For later use, we say that a subset S is *critical* at $\vec{z}$ if $\sum_{j\in S} z_j = r(S)$ and *subcritical* at $\vec{z}$ if $\sum_{j\in S} z_j < r(S)$.

It may seem that we are being inconsistent in our notation, as we have already used K_A to denote a certain convex hull in our description of Barthe's results. We shall show below that in fact the two sets coincide. For present puroses, this is not important, and the definition of K_A shall be the one made just above.

What we have just seen shows that $\vec{p} \in K_A$ is a necessary condidtion for $D(\vec{p}) < \infty$, or even $D_{\mathcal{G}}(\vec{p}) < \infty$. It turns out that it is also sufficient.

Theorem 4.2 (Barthe) *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be any a spanning set of vectors in $\mathbb{R}^M$. Then $D_{\mathcal{G}}(\vec{p}) < \infty$ if and only if $\vec{p} \in K_A$. Moreover, if $\vec{z} \in K_A^\circ$, then the supremum is attained in the Gaussian variational problem (1.3).*

Barthe's proof is based on the convex hull description of K_A, as mentioned above. At the end of this section we give an alternate proof, and show directly that Barthe's convex hull definition of K_A yields the same set as does our definition. First, we deduce the Brascamp–Lieb Theorem from Theorems 3.1, 4.2 and Lemma 4.1.

Theorem 4.3 (Brascamp–Lieb) *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be any a spanning set of vectors in $\mathbb{R}^M$. Then for all $\vec{p}$, $D_{\mathcal{G}}(\vec{p}) = D(\vec{p})$.*

Proof: If $\vec{p} \in K_A^\circ$, everything is clear. By Theorem 4.2, the Gaussian problem has optimizers, and then by Theorem 3.1, $D_{\mathcal{G}}(\vec{p}) = D(\vec{p})$.

Therefore, suppose that $\vec{p} \in K_A$, but not K_A°. Then there exists a non–empty proper subset S of the indices that is critical; i.e, such that $\sum_{j\in S} 1/p_j = r(S)$. We further take S to have the *least cardinality* among all such sets.

To apply the identity (4.2), consider

$$D_S = \sup \left\{ \frac{\int_{\mathbb{R}^r} \prod_{j\in S} f_j(\vec{b}_j \cdot y) \mathrm{d}^r y}{\prod_{j\in S} \|f_j\|_{p_j}} \ : \ f_j \in L^{p_j}(\mathbb{R}) \qquad j = 1, 2, \ldots, N \right\} \tag{4.7}$$

and

$$D_{S^c} = \sup \left\{ \frac{\int_{\mathbb{R}^{M-r}} \prod_{j\in S^c} f_j(\vec{c}_j \cdot z) \mathrm{d}^{M-r} z}{\prod_{j\in S^c} \|f_j\|_{p_j}} \ : \ f_j \in L^{p_j}(\mathbb{R}) \qquad j = 1, 2, \ldots, N \right\} . \tag{4.8}$$

Here, as in (4.2), $r = r(S)$. Clearly, (4.2) yields the bound $D(\vec{p}) \le D_S D_{S^c}$.

To obtain the opposite inequality, note that the the scaling identity $\sum_{j\in S}(1/p_j) = r$ is satisfied, and by the choice of a critical set of minimal cardinality, there are *no* critical subsets of S for the variational problem of computing D_S.

Therefore, there is a solution of the Euler–Lagrange equation (3.12) for (4.7), and hence it has Gaussian maximizers. From (3.20) we see that we can take these maximizers g_j so that $\prod_{j\in S} f_j(\vec{b}_j \cdot y)$ is an arbitrarily good approximation of D_S times a Dirac mass at the

origin. This will eliminate the terms involving y in the second integral in (4.2). Hence for any $\epsilon > 0$, one can choose the functions f_j, $j \in S$, to be Gaussian and have

$$\frac{\int_{I\!R^r} \prod_{j=1}^N f_j(\vec{b}_j \cdot y) \mathrm{d}^r y}{\prod_{j=1}^N \|f_j\|_{p_j}} \geq D_S \left(\frac{\int_{I\!R^{M-r}} \prod_{j \in S^c} f_j(\vec{c}_j \cdot z) \mathrm{d}^{M-r} z}{\prod_{j \in S^c} \|f_j\|_{p_j}} \right) - \epsilon \ .$$

We are now reduced to proving that the variational problem for D_{S^c} has Gaussian maximizers. If there are no critical subsets of S^c for this problem, we are done by Theorem 4.1. Otherwise, "peel off" another critical subset. This procedure clearly reduces the cardinality of S^c each time, and hence it terminates with a full set of Gaussian trial functions that come arbitrarily close to the supremum. This yields the identity $D(\vec{p}) = D_S D_{S^c}$ and completes the proof of Theorem 4.3. ■

The remainder of this section is devoted to the proof of Theorem 4.2. We first show that the two definitions of K_A do indeed define the same set.

Theorem 4.4 *For any spanning set of vectors $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$, K_A is a the convex hull of the vectors $\vec{z}$ whose entries are either zero or one, and for which the set $\{\vec{a}_j \ : \ z_j = 1 \}$ is a basis of $I\!R^M$.*

First we prove a lemma.

Lemma 4.5 *Consider any spanning set of vectors $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$, and any $\vec{z}$ in K_A. Let T be any non–empty subset of the indices $\{1, 2, \ldots, N\}$. If there is any set S of indices containing T that is critical at $\vec{z}$, then there is a least such set S_0: That is, there is a set S_0 containing T that is critical at $\vec{z}$, such that if $\tilde{S}$ is any other set containing T that is critical at $\vec{z}$, then $S_0 \subset \tilde{S}$. .*

Proof: Without loss of generality, we may suppose that there is a set of indices containing T that is critical at $\vec{z}$. Let S be such a set of least cardinality, and let $\tilde{S}$ be any other set of indices containing T that is critical at $\vec{z}$. Let V be the span of $\{\vec{a}_j \ : \ j \in S\}$, and let W be the span of $\{\vec{a}_j \ : \ j \in \tilde{S}\}$. Clearly,

$$\{\vec{a}_j \ : \ j \in S\} \cap \{\vec{a}_j \ : \ j \in \tilde{S}\} \subset V \cap W$$

and

$$\{\vec{a}_j \ : \ j \in S\} \cup \{\vec{a}_j \ : \ j \in \tilde{S}\} \subset V \cup W \ .$$

From the relation $\dim(V \cap W) + \dim(V \cup W) = \dim(V) + \dim(W)$ it then follows that

$$r(S \cap \tilde{S}) + r(S \cup \tilde{S}) \leq r(S) + r(\tilde{S}) \ . \tag{4.9}$$

Since both S and $\tilde{S}$ are critical at z, $r(S) + r(\tilde{S}) = \sum_{j \in S} z_j + \sum_{j \in \tilde{S}} z_j$. Since $\vec{z} \in K_A$, $\sum_{j \in S \cup \tilde{S}} z_j \leq r(S \cup \tilde{S})$ thus from (4.9)

$$\sum_{j \in S \cup \tilde{S}} z_j + r(S \cap \tilde{S}) \leq \sum_{j \in S} z_j + \sum_{j \in \tilde{S}} z_j \ .$$

This implies that $r(S \cap \tilde{S}) \leq \sum_{j \in S \cap \tilde{S}} z_j$ and since T is non–empty and is a subset of both S and $\tilde{S}$, $S \cap \tilde{S}$ is not empty. Hence $S \cap \tilde{S}$ is critical at z. If $S \cap \tilde{S}$ were a proper subset of S, then we would have found a critical subset of strictly smaller cardinality, contrary to the assumption on S. Hence $S = S \cap \tilde{S}$, and so $S \subset \tilde{S}$. The set S is the set S_0 that we seek. ■

Proof of Theorem 4.4: Suppose that $\vec{z} \in K_A$, and for some k, $0 < z_k < 1$. We shall show that in this case, $\vec{z}$ is not extreme.

First, consider the case in which no critical set contains any indices j for which $0 < z_j < 1$.

Since $\sum_{j=1}^{N} z_j = M$, which is an integer, it must be the case that for some $\ell \neq k$, $0 < z_\ell < 1$. Since neither k nor ℓ belongs to any critical set, increase (resp. decrease) z_k a little, while decreasing (resp. increasing) z_ℓ a little in such a way that $z_k + z_\ell$ is constant, and the increases do not produce any supercritical sets. Clearly in this case, $\vec{z}$ is not extreme.

Second, if there are critical sets containing indices j for which $0 < z_j < 1$, choose one, S, of least cardinality. Since S is critical, $\sum_{j \in S} z_j$ is an integer, and there must be two indices k and ℓ in S such that $0 < z_k, z_\ell < 1$. By the lemma and the definition of S, S is the smallest critical set containing either k or ℓ.

Clearly, we can increase z_k a little bit, and decrease z_ℓ a little bit without affecting $z_k + z_\ell$, and hence without affecting $\sum_{j \in S} z_j$. Moreover, the increase in z_k does not increase the value of $\sum_{j \in \tilde{S}} z_j$ for any other critical set $\tilde{S}$ that contains k. This is because $S \subset \tilde{S}$ by Lemma 4.5 and the definition of S, and hence $\tilde{S}$ also contains ℓ. ■

Proof of Theorem 4.2: Let $\phi_A(t_1, t_2, \ldots, t_N) = \ln\left(\det(Ae^T A^t)\right)$. The function ϕ_A was shown to be convex on $I\!R^N$ by Brascamp and Lieb. Let ϕ_A^* denote its Legendre transform:

$$\phi_A^*(z_1, z_2, \ldots, z_N) = \sup_{t_1, t_2, \ldots, t_N} \left\{ \sum_{j=1} z_j t_j - \phi_A(t_1, t_2, \ldots, t_N) \right\} \tag{4.10}$$

By (3.10), determining the set of vectors $\vec{p}$ for which $D_{\mathcal{G}}(\vec{p}) < \infty$ is the same as determining the set of vectors $\vec{z}$ for which $\phi_A^*(\vec{z}) < \infty$.

Next, recall a formula of Brascamp and Lieb, which can be deduced from the Cauchy–Binet formula:

$$\det\left(Ae^T A^t\right) = \sum_{|S|=M} t_S \det\left(A_S A_S^t\right) \ , \tag{4.11}$$

where $t_S = \exp\left(\sum_{j \in S} t_j\right)$. Here, we use the following notation: If $S = \{j_1, j_2, \ldots, j_k\}$, A_S denotes the $M \times k$ matrix

$$A_S = [\vec{a}_{j_1}, \vec{a}_{j_2}, \ldots, \vec{a}_{j_k}] \tag{4.12}$$

As shown in [5], the convexity of $\phi_A(\vec{t})$ follows by differentiating $\phi(\vec{t}+r\vec{v})$ twice with respect to r using the Schwarz inequality. Here $\vec{v}$ is an arbitrary fixed vector.

Having made these remarks, we first show that $\phi^*(\vec{z}) = \infty$ unless $\sum_{j=1}^N z_j = M$, $0 \leq z_j \leq 1$ for each j.

For any constant c and any $\vec{t}$ in $\mathbb{R}^N$, let $\vec{t}_c$ denote the vector in $\mathbb{R}^N$ whose jth component is $t_j + c$. From the definition (3.9), it follows that $\phi_A(\vec{t}_c) = Mc + \phi_A(\vec{t})$. Therefore,

$$\vec{z}\cdot\vec{t}_c - \phi_A(\vec{t}_c) = \left(\sum_{j=1}^N z_j - M\right) c + \vec{z}\cdot\vec{t} - \phi_A(\vec{t}) \tag{4.13}$$

so that the domain of ϕ_A^* lies in the hyperplane $\sum_{j=1}^N z_j = M$. Further, it follows from (3.11) that $0 \leq \dfrac{\partial}{\partial t_j}\phi_A(\vec{t}) \leq 1$ since this quantity is is the jth diagonal entry of an orthogonal projection. Hence, every $\vec{z}$ in K_A is such that $0 \leq z_j \leq 1$ for each j.

Recall the terminology that a subset S is *critical* at $\vec{z}$ if $\sum_{j\in S} z_j = r(S)$ and *subcritical* at $\vec{z}$ if $\sum_{j\in S} z_j < r(S)$.

We now show that $\phi_A^*(\vec{z}) < \infty$ if $\vec{z} \in K_A^\circ$. First, note that if $\sum_{j=1}^N z_j = M$, then (4.13) reduces to

$$\vec{z}\cdot\vec{t}_c - \phi_A(\vec{t}_c) = \vec{z}\cdot\vec{t} - \phi_A(\vec{t})\ . \tag{4.14}$$

Hence in (4.10), we may restrict our focus to vectors $\vec{t}$ satisfying $\min_{j=1}^N t_j = 0$.

For any $\vec{t} = (t_1, \ldots, t_N)$, let $\vec{t}^* = (t_1^*, \ldots, t_N^*)$ be its decreasing rearrangement. By the invariance noted above, we may assume that $t_N^* = 0$. Let π be any permutation so that

$$t_j^* = t_{\pi(j)} \qquad \text{for all} \quad 1 \leq j \leq N\ .$$

Let $\tilde{S}$ be the indices of the pivotal columns in $\pi(A) = [\vec{a}_\pi(1), \vec{a}_\pi(2), \ldots, \vec{a}_\pi(N)]$. That is, the columns of $A_{\tilde{S}}$ are the columns in $\pi(A)$ that are not in the span of the columns to their left in $\pi(A)$. Since the dimension of the space spanned by the vectors $\vec{a}_1, \ldots \vec{a}_N$ is M we have that $|\tilde{S}| = M$. By monotonicty of the logarithm and (4.11),

$$\phi_A(t_1, t_2, \ldots, t_N) = \ln(\det(Ae^T A^t)) \geq \sum_{j\in\tilde{S}} t_j + \ln(\det(A_{\tilde{S}} A_{\tilde{S}}^t)\ , \tag{4.15}$$

and hence it suffices to find a lower bound on

$$\sum_{j\in\tilde{S}} t_j - \sum_{j=1}^N z_j t_j\ . \tag{4.16}$$

Setting $a_k = 1$ if $\pi^{-1}(k) \in S$, and $a_k = 0$ otherwise and $b_k = z_{\pi(k)}$, (4.16) can be written as

$$\sum_{k=1}^{N} (a_k - b_k) t_k^* \ .$$

The point about this notation is that the vector $(a_1, \ldots, a_M, 0, \ldots, 0)$ which has N elements strictly majorizes the vector $(b_1, \ldots b_N)$, i.e.,

$$\sum_{j=1}^{k} a_k > \sum_{j=1}^{k} b_k, k = 1, \ldots, N-1 \ , \tag{4.17}$$

and

$$\sum_{j=1}^{N} a_k = \sum_{j=1}^{N} b_k = M \ . \tag{4.18}$$

The equation (4.18) follows from the definition of the a_k's and b_k's, the fact that $|\tilde{S}| = M$ and the fact that $\sum_{j=1}^{N} z_j = M$. The equation (4.17) follows from the definition of the a_k's and b_k's and the fact that $\vec{z} \in K_A^\circ$, i.e., for every proper subset S of $\{1, \ldots, N\}$, $\sum_{j \in S} z_j < r(S)$.

Summing by parts, using $t_N^* = 0$ and $\sum_{k=1}^{N} a_k = \sum_{k=1}^{N} b_k$,

$$\sum_{k=1}^{N} (a_k - b_k) t_k^* = \sum_{k=1}^{N-1} \left(\sum_{j=1}^{k} (a_j - b_j) \right) (t_k^* - t_{k+1}^*) \geq c t_1^* \ ,$$

where $c = \min_{1 \leq k \leq N-1} \sum_{j=1}^{k} (a_j - b_j) > 0$.

Hence

$$\sum_{k \in S} t_k - \sum_{k=1}^{N} z_k t_k \geq c \max_k (t_k) \ ,$$

which, together with the bound (4.15), yields

$$\vec{z} \cdot \vec{t} - \phi_A(\vec{t}) \leq -\ln(\det(A_{\tilde{S}} A_{\tilde{S}}^t) - c \max_k (t_k) \ .$$

Therefore, as any of the variables $t_1, \ldots, t_{N-1}$ tend to infinity (recall that without loss t_N can be chosen to be zero), $\vec{z} \cdot \vec{t} - \phi_A(t_1, \ldots, t_N)$ tends to $-\infty$, and so $\phi_A^*(\vec{z}) < \infty$. By the convexity of ϕ_A, proved by Brascamp and Lieb, the supremum in (4.10) is attained in this case.

It remains to show that $\phi_A^*(\vec{z}) < \infty$ for all $\vec{z}$ in K_A. This is an easy consequence of Theorem 4.4. Suppose that $\vec{p}$ is one of the verticies of K_A. If $p_j = \infty$, we may as well

replace f_j by 1 in (1.2). Therefore, there are effectively only M vectors and functions. Letting S denote the set of indices for which $p_j = 1$, we have the identity

$$\int_{I\!R^M} \prod_{j\in S}^{M} f_j(\vec{a}_j \cdot x) \mathrm{d}^M x = (\det(A_S A_S^t))^{1/2} \prod_{j\in S}^{M} \int_{I\!R} f_j(y) \mathrm{d}y \ . \tag{4.19}$$

which gives us $D(\vec{p}) = (\det(A_S A_S^t))^{1/2}$. This is finite, and since $D(\vec{p})$ is convex and finite at the vertices of K_A, it is finite throughout K_A. ■

5: Determination of the optimizers

A partial solution to the problem of determining all maximizers, when they exist, for (1.2) was obtained in [5] where it is proved that in the case $M = 2$ and $N = 3$, which gives the classical Young's inequality, the only non negative maximizers of the ratio in (1.2) are certain specific Gaussian functions. The method of proof extends to more general cases involving $M + 1$ functions in $I\!R^M$, but not to values of $N > M + 1$.

Under the additional assumption that there exists an optimizer to the Gaussian variational problem (1.3), a full determination of the non negative optimizers was obtained by Barthe [3]. He conjectures that when there is no optimizer to the Gaussian problem, there is no optimizer for the general problem (1.2). Here we give a proof of Barthe's theorem, and of his conjecture. We also determine the form of all of the complex valued optimizers.

Before we begin, note a restriction that we may impose on $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ without loss of generality: We may assume that if any one vector is deleted from $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$, then what remains still spans $I\!R^M$. The point is that when $\vec{a}_1$ is necessary for the whole set to span, there is a change of coordinates under which

$$\int_{I\!R^M} \prod_{j=1}^{N} f_j(\vec{a}_j \cdot x) \mathrm{d}^N x = \frac{1}{|\vec{u}_1 \cdot \vec{a}_1|} \int_{I\!R} f_1(z_1) \mathrm{d}z_1 \left(\prod_{j=2}^{N} \int_{I\!R^{M-1}} f(\vec{b}_j \cdot \vec{w}) \mathrm{d}^{M-1} w \right) \ , \tag{5.1}$$

for some vectors $\vec{b}_2, \ldots, \vec{b}_N$ in $I\!R^{M-1}$. (The calculation in (5.1) is carried out at the end of the Appendix.) This reduces the analysis of (3.1) to an integral of the same type, but with one factor and one dimension fewer. It also shows that in this case, we must have $p_1 = 1$ to obtain a finite constant D. Also it is clear in this case that the optimizers need not be Gaussian, since f_1 can be any integrable function without affecting the value of the ratio in (1.2).

We therefore make the following definition:

Definition Given a spanning set $S = \{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ of vectors in $I\!R^M$, we say that $\vec{a}_j$ is *essential* in case $S \backslash \{\vec{a}_j\}$ does not span $I\!R^M$, and we say that S is *properly redundant* in case no vector in S is essential, and moreover, no two vectors in S are proportional.

We can always apply the reduction argument given just above to eliminate any essential vectors. Notice that if $N = M$, every vector is essential, and in fact, we have the identity

$$\int_{I\!R^M} \prod_{j=1}^{N} f_j(\vec{a}_j \cdot x) \mathrm{d}^N x = (\det(AA^t))^{1/2} \prod_{j=1}^{M} \int_{I\!R} f_j(y) \mathrm{d}y \ . \tag{5.2}$$

For this reason, we are interested mainly in $N > M$.

We also see right away from (5.2) that if $\vec{p}$ is a vertex of K_A, so that M of the L^p indices are 1, and $N - M$ are ∞, then we get maximizers in (1.2) if and only if we take each of the L^∞ functions to be constant, and there is no restriction on the L^1 functions. Hence for $\vec{p}$ a vertex of K_A, the maximizers are far from unique, and need not be Gaussian.

Lemma 5.1 Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ span $I\!R^M$. Let $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N]$ be the $M \times N$ matrix whose j-th column is $\vec{a}_j$. Let P be the orthogonal projection in $I\!R^N$ onto the image of A^t. Then $\vec{a}_j$ is essential if and only if $P_{j,j} = 1$.

Proof: By definition, $\vec{a}_j$ is essential if and only if there do not exist numbers $u_1, u_2, \ldots, u_N$ such that

$$u_j \neq 0 \qquad \text{and} \qquad \sum_{k=1}^{N} u_k \vec{a}_k = 0 \ . \tag{5.3}$$

Let $\vec{u}$ be the vector in $I\!R^N$ whose kth entry is u_k. Then $\sum_{k=1}^{N} u_k \vec{a}_k = 0$ is exactly the condition for $\vec{u}$ to belong to the kernel of A. Hence, $\vec{a}_j$ is essential if and only if $u_j = 0$ for each vector $\vec{u}$ in the kernel of A. Let $\vec{e}_j$ denote the jth standard basis vector in $I\!R^N$, so that $u_j = \vec{e}_j \cdot \vec{u}$. Then, since the image of A^t is the orthogonal complement of the kernel of A, we have that $\vec{a}_j$ is essential if and only if $\vec{e}_j$ belongs to the image of A^t. Clearly, this is the case if and only if $P_{j,j} = 1$. ■

Theorem 5.2 (Barthe) *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be any properly redundant spanning set, and let $\vec{p} \in K_A^\circ$. Then the solution S of the Euler–Lagrange equations (3.12) is unique up to a constant multiple. Moreover, non-negative functions $f_1, \ldots, f_N$ satisfy*

$$\int_{I\!R^M} \prod_{j=1}^{N} f_j(\vec{a}_j \cdot x) \mathrm{d}^M x = D(\vec{p}) \prod_{j=1}^{N} \|f_j\|_{p_j} \tag{5.4}$$

if and only if there is a number $c > 0$ and a vector $\vec{b} \in \mathrm{Img}(A^t)$ so that for each j, $f_j(y)$ is a multiple of

$$\exp\left(-\frac{c}{2}\left(s_j^2 (y - b_j)^2\right)\right) \ . \tag{5.5}$$

Proof: Recall the proof of Theorem 3.1. Fix any $t > 0$, and let $h_j(y)$ denote $\ln(f_j(t,y))$. Note that each $f_j(t,y)$ is smooth and strictly positive, so that each h_j is smooth. Then by (3.16) we must have

$$\int_{I\!R^M} \left[\sum_{i,j} h'(\vec{a}_i \cdot x) Q_{i,j} h'(\vec{a}_j \cdot x) \right] F(x) \mathrm{d}^M x = 0 \ , \tag{5.6}$$

where $F(x) = \prod_{j=1}^N f_j(\vec{a}_j \cdot x)$ and Q is given by (3.17): $Q = S^{-1}(I-P)S^{-1}$ where P is the orthogonal projection onto the image of $(AS)^t = SA^t$.

Step 1: (Each h_j is a quadratic polynomial) Since F is strictly positive, it follows from (5.6) that the vector

$$\vec{v}(x) = \begin{bmatrix} h_1'(\vec{a}_1 \cdot x) \\ h_2'(\vec{a}_2 \cdot x) \\ \vdots \\ h_N'(\vec{a}_N \cdot x) \end{bmatrix} \ , \tag{5.7}$$

is such that $S^{-1}\vec{v}(x)$ lies in $\mathrm{Img}(SA^t)$ for every x. This means that if $\vec{u}$ is any vector in the kernel of AS, then $\vec{u} \cdot S^{-1}\vec{v}(x) = 0$ for all x.

For any vector $\vec{u}$ in the kernel of AS, we define the function $\phi(x)$ by

$$\phi(x) = \vec{u} \cdot S^{-1}\vec{v}(x) = \sum_{j=1}^N u_j s_j^{-1} h_j'(\vec{a}_j \cdot x) \ .$$

Since ϕ vanishes identically, $0 = \nabla\phi(x) = \sum_{j=1}^N u_j s_j^{-1} h_j''(\vec{a}_j \cdot x)\vec{a}_j$. This means that for each x, the vector $\vec{w}(x)$ defined by

$$\vec{w}(x) = \begin{bmatrix} s_1^{-1} h_1''(\vec{a}_1 \cdot x) u_1 \\ s_2^{-1} h_2''(\vec{a}_2 \cdot x) u_2 \\ \vdots \\ s_N^{-1} h_N''(\vec{a}_N \cdot x) u_N \end{bmatrix} \ , \tag{5.8}$$

lies in the kernel of A.

We shall first show that $h_1''(\vec{a}_1 \cdot x)$ is constant. To do this, write $\vec{a}_1$ as a linear combination of the other vectors $\vec{a}_j$: $\vec{a}_1 = \sum_{j=2}^N \alpha_j \vec{a}_j$. This is possible since $\vec{a}_1$ is not essential. There may be many ways of doing this, but we can always choose one such that a minimal number of the α's are non zero, which we do. Suppose that there are exactly k values of j, $j_1, j_2, \ldots, j_k$ for which $\alpha_j \neq 0$

The vector $\vec{u} = S^{-1} \begin{bmatrix} 1 \\ -\alpha_2 \\ \vdots \\ -\alpha_N \end{bmatrix}$ belongs to the kernel of AS. We use this vector $\vec{u}$ in (5.8) to define $\vec{w}(x)$.

Now let $\vec{y}$ be any vector in $\mathbb{R}^M$ that is orthogonal to $\vec{a}_{j_k}$, but not orthogonal to $\vec{a}_1$. Since $w(x)$ lies in the kernel of A for every x, so does the vector we get when we differentiate each component in the $\vec{y}$ direction. That is, for each x,

$$\begin{bmatrix} (\vec{y}\cdot\vec{a}_1)s_1^{-1}h_1'''(\vec{a}_1\cdot x)u_1 \\ (\vec{y}\cdot\vec{a}_2)s_2^{-1}h_2'''(\vec{a}_2\cdot x)u_2 \\ \vdots \\ (\vec{y}\cdot\vec{a}_2)s_N^{-1}h_N'''(\vec{a}_N\cdot x)u_N \end{bmatrix}$$

lies in the kernel of A. The j_kth component of this vector vanishes identically since $\vec{y}\cdot\vec{a}_{j_k} = 0$. The $N-k$ other entries for which $u_j = 0$ also vanish identically. This means that for each x, the above vector lies in the kernel of A, and has no more than $k-1$ non zero entries. By assumption there is no vector in the kernel of A whose first component is non zero and that has fewer than k non zero entries. Hence the first component must be zero. Since $\vec{y}\cdot\vec{a}_1 \neq 0$, and $u_1 \neq 0$, this means $h_1'''(\vec{a}_1\cdot x) = 0$, and proves that h_1'' is constant.

The argument may now be repeated for each j, and we learn at this point that each h_j is a quadratic function.

Step 2: (Determination of h_j'') Let c_j denote the value of h_j''. Then, from (5.8), for any vector $\vec{u}$ in the kernel of AS, the vector whose jth entry is $s_j^{-1}c_ju_j$ belongs to $\mathrm{Ker}(A)$. Since $\vec{u}$ is in the kernel of AS if and only if $S\vec{u}$ is in the kernel of A, we see that $s_j^{-1}c_j$ must be a constant multiple of s_j. In other words, for some constant c, we have

$$s_1^{-2}h_1''(\vec{a}_1\cdot x) = s_2^{-2}h_2''(\vec{a}_2\cdot x) = \cdots = s_N^{-2}h_N''(\vec{a}_N\cdot x) = -c \tag{5.9}$$

This of course means that for each j, there are constants a_j and b_j so that

$$h_j(y) = -\frac{c}{2}s_j^2\,(y-b_j)^2 + a_j\ , \tag{5.10}$$

which means that

$$f_j(y) = \exp\left(-\frac{c}{2}s_j^2\,(y-b_j)^2 + a_j\right)\ . \tag{5.11}$$

Evidently, $c > 0$.

Step 3: (Determination of b_j) Let $\vec{b}$ denote the vector $\vec{b}$ in $\mathbb{R}^N$ whose jth component is b_j. From (5.10) and the definition of $\vec{v}(x)$, we see that $\vec{v}(0) = -cS^2\vec{b}$. We have seen that $S^{-1}\vec{v}(0)$ lies in $\mathrm{Img}(SA^t)$, and so $\vec{b}$ lies in $\mathrm{Img}(A^t)$.

The constant b_j is the mean of the probability density $f_j^{p_j}(y)/\|f_j\|_{p_j}^{p_j}$, and the mean does not change under the evolution considered here, which commutes with translations. Therefore, we see that $f_j(t,y)$ can have the form specified in (5.11) if and only if it has this form at $t=0$. That is, there is equality in (5.4) if and only if there is a positive constant c, and a vector $\vec{b}$ in the image of A^t so that each f_j has the form specified in (5.11) with b_j being the jth component of $\vec{b}$, and the a_j are arbitrary. ■

Corrolary 5.3 *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be any properly redundant spanning set. Then the function $\phi(t_1, t_2, \ldots, t_N) = Tr\left(\ln\left(Ae^T A^t\right)\right)$ is strictly convex, except along the lines obtained by adding a number c to each t_j.*

Proof: Were this not the case, we would have two solutions S of the Euler-Lagrange that would not be constant multiples of one another. ■

The *strict* convexity was proved by Brascamp and Lieb under the stronger hypothesis that every subset of M vectors chosen from $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ is linearly independent.

Concerning maximizers for $\vec{p}$ on the boundary of K_A, we have already dealt with the vertices – these have plenty of non–Gaussian maximizers, and in the strict sense considered here do not have any Gaussian optimizers: If $p_j = 1$, then f_j may be any non negative L^1 function, and and so may be taken to be Gaussian, while if $p_j = \infty$, then f_j must be constant, and therefore not Gaussian. One could consider constants as degenerate Gaussians, though this would not be entirely consistent with the terminology we have been using in reference to the Gaussian optimization problem. Alternately, one can stipulate that $p_j < \infty$ for all j. Indeed, if $p_j = \infty$, then the corresponding factors involving f_j can be deleted top and bottom in (1.2) without affecting the constant. We may then prove a conjecture of Barthe [3]:

Theorem 5.4 *Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be any properly redundant spanning set, and let $\vec{p} \in K_A \backslash K_A^\circ$ be such that $p_j < \infty$ for all j. Then there may be no optimizers for the Brascamp Lieb inequality, but when there are optimizers, there are Gaussian optimizers. Moreover, there is a constructive procedure for deciding whether or not optimizers exist in any particular case.*

Proof: We again apply the factorization formula (4.2) from Lemma 4.1. As in the proof of Theorem 4.3, let S be a critical set of least cardinality. Such a set exists since $\vec{p} \in K_A \backslash K_A^\circ$. As shown in the proof of Theorem 4.3, $D(\vec{p}) = D_S D_{S^c}$ where D_S and D_{S^c} are defined in (4.7) and (4.8) respectively. Since S was a critical subset of least cardinality, there are no critical subsets for this problem. Hence there are Gaussian optimizers for the variational problem that determined D_S.

Next, suppose that there are no critical sets in the variational problem that determines D_{S^c}. Then this problem has only Gaussian optimizers, unique up to a common scaling and certain translations. However, examining (4.2) we see that for

$$\int_{I\!R^r} \prod_{j\in S} f_j(\vec{b}_j \cdot y) \mathrm{d}^r y \left(\int_{I\!R^{M-r}} \prod_{j\in S^c} f_j(\vec{b}_j \cdot y + \vec{c}_j \cdot z) \mathrm{d}^{M-r} z \right) \tag{5.12}$$

to equal $D_S D_{S^c} \prod_{j=1}^N \|f\|_{p_j}$, it is necessary that the translations in the integral on the right be among those permitted by Theorem 5.2. There is a simple criterion for this: Let A_{S^c} be the matrix obtained by deleting from A the jth collumn whenever $j \in S$. Let U and V be the partial isometries used in the proof of Lemma 4.1, so that if we put $B = A_{S^c}^t U$ and $C = A_{S^c}^t V$, the columns of B (resp. C) are the vectors b_j (resp. c_j) in the second integral.

When optimizers exist, it must be the case that for each y the translation in the second integral is one permitted by Theorem 5.2. Clearly this is the case if and only if $\mathrm{Img}(B) \subset \mathrm{Img}(C)$. Conversely, if this is the case, all of the translation are admissible, and using Gaussian optimizers for D_{S^c} in (5.12), we will have this integral equal to $D_S D_{S^c} \prod_{j=1}^N \|f\|_{p_j}$.

The general case is handled in very much the same way: If there are critical sets in the variational problem that determines D_{S^c}, "peel these off" repeatedly until one gets a problem with no critical subsets, and hence Gaussian optimizers. Now one works ones way back up, checking the compatibility condition $\mathrm{Img}(B) \subset \mathrm{Img}(C)$ each step of the way. If this is ever violated, there are no optimizers. Otherwise, we obtain a set of Gaussian optimizers. ■

One might further hope that the Gaussian functions in Theorem 5.2 are also the only optimizers of Young's inequality in the wider class of complex valued functions. However, this is not the case. The reason is that there exist in general functions $\phi_j(y)$ with

$$e^{i \sum_{j=1}^N \phi_j(\vec{a}_j \cdot \vec{x})} = 1 \ , \tag{5.13}$$

and thus if $f_1, \dots, f_N$ is any set of non-negative optimizers, then $e^{\phi_1} f_1, \dots, e^{\phi_N} f_N$ is a set of complex optimizers. Here are some examples. Any three vectors in $\mathbb{R}^2$ are linearly dependent, i.e., there is a relation $\sum_{j=1}^3 \alpha_j \vec{a}_j = 0$. hence with $\phi_j(y) = \alpha_j y$ (5.13) holds. With four vectors there are more possibilities. E.g., pick $\vec{a}_1 = \epsilon_1, \vec{a}_2 = \vec{e}_2, \vec{a}_3 = (\vec{e}_1 + \vec{e}_2)/\sqrt{2}$ and $\vec{a}_4 = (\vec{e}_1 - \vec{e}_2)/\sqrt{2}$. then the function $\phi_1(y) = \phi_2(y) = -y^2$, $\phi_3(y) = \phi_4(y) = y^2$ again satsify (5.13). Hence there are non-trivial complex valued optimizers. *

In general, let $\{f_1, f_2, \dots, f_N\}$ be any set of optimizers. Define functions z_j by

$$z_j(y) = f_j(y)/|f_j(y)|$$

where $f_j(y) \neq 0$, and $z_j(y) = 1$ otherwise. These functions take values in the unit circle in the complex plane. In order to have equality in the generalized Young inequality, it is necessary that

$$\prod_{j=1}^N z_j(\vec{a}_j \cdot \vec{x}) = 1 \tag{5.14}$$

* The possibility of complex optimizers of this type for Young's inequality was pointed out to Brascamp and Lieb by J. Fournier; see a note added in proof at the end of their paper.

almost everywhere.

Theorem 5.5 *Let $\{\vec{a}_1, \ldots, \vec{a}_N\}$ be any set of vectors spanning $\mathbb{R}^M$ such that no two vectors are multiples of one another. Let z_j, $j = 1, \ldots, N$, be any j measurable functions from $\mathbb{R}$ to the unit circle in the complex plane such that (5.14) holds almost everywhere. Then for each j,*

$$z_j(y) = e^{i\phi_j(y)}$$

where ϕ_j is a polynomial of degree at most $N - M$.

We first prove a lemma:

Lemma 5.6 *Let z be a function from $\mathbb{R}$ to the unit circle in the complex plane, and let n be any positive integer. Suppose that z has the following property:*

$$\frac{z(x+y)}{z(x)} = e^{i\psi(x,y)}$$

where $\psi(x, y)$ is a polynomial of degree $n - 1$ in x with coefficients that a measurable functions of y. Then $z(x) = e^{i\phi(x)}$ where ϕ is a polynomial of degree n.

Proof: Before beginning, notice that the modulus of z is constant, and non zero. Hence z is never zero.

Consider first the case $n = 1$. Writing $w(y) = e^{\psi(y)}$, since there is no x dependence in this case,

$$z(x+y) = z(x)w(y) \ .$$

Now let ρ be any smooth compactly supported function on $\mathbb{R}$. Then

$$\int z(x+y)\rho(y)\mathrm{d}y = z(x)\int w(y)\rho(y)\mathrm{d}y \ .$$

Since the restriction of w to any interval is a non–zero function in L^2 on that interval, and since smooth, compactly supported functions are dense in this L^2 space, we can choose ρ so that $\int w(y)\rho(y)\mathrm{d}y = c \neq 0$. Then we have

$$z(x) = \frac{1}{c}\int z(x+y)\rho(y)\mathrm{d}y \ .$$

This shows that z is smooth. In particular, once we chose a branch of the logarithm for $z(0)$, there is just one way to choose the logarithm of $z(x)$ so that it is continuous, and then of course it is smooth. Hence there is a smooth real function ϕ so that $z(x) = e^{i\phi(x)}$, and

$$\phi(x+y) = \phi(x) + \psi(y) \ .$$

Evidently, ψ is also smooth. Applying $\partial^2/\partial x\partial y$ to both sides, we learn that ϕ'' vanishes identically, and so ϕ is a polynomial of first degree.

Now suppose that $n \geq 2$. Here the argument is similar, but requires one more step. We first write

$$z(x+y) = z(x)e^{i\psi(x,y)} \ .$$

Pick any x_0, and choose a smooth, compactly supported function ρ as above so that for this x_0,

$$\int e^{i\psi(x_0,y)}\rho(y)\mathrm{d}y \neq 0 \ .$$

Now, no matter how large the coefficients of the polynomial $\psi(x,y)$ may be at certain y in the support of ρ, the function

$$x \mapsto \int e^{i\psi(x,y)}\rho(y)\mathrm{d}y$$

is continuous in x by the Dominated Convergence Theorem.

We conclude that $c(x) = \int e^{i\psi(x_0,y)}\rho(y)\mathrm{d}y$ is continuous and non–zero on a neighborhood of x_0. Hence

$$z(x) = \frac{1}{c(x)}\int z(x+y)\rho(y)\mathrm{d}y$$

is continuous on a neighborhood of x_0. Since x_0 is arbitrary, z is continuous.

It now follows that $e^{i\psi(x,y)}$ is continuous in both x and y. Therefore, the coefficients are uniformly bounded functions of y in any compact interval. This means that all of the partial derivatives in x of $e^{i\psi(x,y)}$ are integrable and continuous, and so the function $c(x) = \int e^{i\psi(x,y)}\rho(y)\mathrm{d}y$ that we defined above is not only continuous, it is actually smooth in x. It now follows that z is smooth, and as before we write $z = e^{i\phi}$, and have

$$\phi(x+y) = \phi(x) + \psi(x,y) \ .$$

Taking n derivatives in x, and using the hypothesis that $\psi(x,y)$ has degree $n-1$ in x, we see that the nth derivative of ϕ is constant. Hence ϕ is a polynomial of degree n. ■

It is of course well known that if ϕ and ψ are two measurable functions on $\mathbb{R}$ such that

$$\phi(x+y) = \phi(x) + \psi(y)$$

then both ϕ and ψ are first degree polynomials. Lemma 5.6 generalizes this in several respects. It seems likely that it may be known, but we cannot find any reference for it.

Proof of Theorem 5.5: We can easily eliminate any essential vectors from $\{\vec{a}_1, \ldots, \vec{a}_N\}$: If $\vec{a}_j$ is essential, it is clear that z_j is constant. Hence we may assume that $\{\vec{a}_1, \ldots, \vec{a}_N\}$ is properly spanning.

It suffices by symmetry to show that z_1 has the specified form. Choose a basis for $\mathbb{R}^M$ from $\{\vec{a}_1, \ldots, \vec{a}_N\}$ that contains $\vec{a}_1$. After renumbering, we may assume this is

$\{\vec{a}_1, \ldots, \vec{a}_M\}$. Let $\vec{b}_1$ be unit vector that is orthogonal to the span of $\{\vec{a}_2, \ldots, \vec{a}_M\}$, and scaled so that $\vec{b}_1 \cdot \vec{a}_1 = 1$.

Now, for any y_1 in $\mathbb{R}$, translate the identity (5.14) by replacing x with $x + y_1\vec{b}_1$. Since $\vec{b}_1$ is orthogonal to $\vec{a}_j$ for $2 \leq j \leq M$, the corresponding factors are unaffected by translation, and hence

$$z_1((\vec{a}_1 \cdot \vec{x}) + y_1)(\prod_{j=M+1}^{N} z_j(\vec{a}_j \cdot (\vec{x} + y_1\vec{b}_1)) = z_1(\vec{a}_1 \cdot \vec{x}) \prod_{j=M+1}^{N} z_j(\vec{a}_j \cdot \vec{x}) \ .$$

Let T_y be the operator

$$T_y(z)(x) = \frac{z(x+y)}{z(x)} \ .$$

Then defining

$$z_1^{(1)}(t; y_1) = \frac{T_{y_1} z_1}{z_1}(t) \ ,$$

and defining $w_j = z_j(\vec{a}_j \cdot (\vec{x} + y_1\vec{b}_1))/z_j(\vec{a}_j \cdot \vec{x})$ for $j \geq M+1$,

$$z_1^{(1)}(\vec{a}_1 \cdot x; y_1) \prod_{j=M+1}^{N} w_j(\vec{a}_j \cdot \vec{x}) = 1 \ . \tag{5.15}$$

This is of the same form as (5.14), but with fewer functions.

Next, choose $\vec{b}_2$ so that $\vec{b}_2 \cdot \vec{a}_{M+1} = 0$ (if it wasn't the case already that $\vec{b}_1 \cdot \vec{a}_{M+1} = 0$), but $\vec{b}_2 \cdot \vec{a}_1 = 1$. Making the same sort of translation in (5.15), but this time by $y_2\vec{b}_2$, we eliminate the second factor by dividing through, so that the first factor becomes

$$z_1^{(2)}(t; y_1, y_2) = \frac{T_{y_2} z_1^{(1)}}{z_1^{(1)}}(t; y_1) \ .$$

Proceeding in this way, we eventually learn that for some $k < M - N$,

$$\frac{T_{y_{k+1}} z_1^{(k)}}{z_1^{(k)}}(t; y_1, \ldots, y_k)$$

is independent of t.

By Lemma 5.6, it follows that $z_1^{(k)}(t; y_1, \ldots, y_k) = e^{i\phi(t; y_1, \ldots, y_k)}$ where $\phi(t; y_1, \ldots, y_k)$ is a first degree polynomial in t with coefficients that are measurable in $y_1, \ldots, y_k$. But by definition,

$$z_1^{(k)}(t; y_1, y_2) = \frac{T_{y_k} z_1^{(k-1)}}{z_1^{(k-1)}}(t; y_1, \ldots, y_{k-1}) \ .$$

Applying Lemma 5.6 again, we learn the form of $z_1^{(k-1)}$. Proceeding in this way, we learn the form of z_1. ■

Once one knows that the possible phase functions are polynomials of limited degree, it is a problem in linear algebra to determine them explicitly for any particular set of vectors $\{\vec{a}_1, \ldots, \vec{a}_N\}$.

6: The best best constant

Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be a properly redundant spanning set in $\mathbb{R}^M$, and let P denote the orthogonal projection onto the image of A^t. Notice that $Tr(P) = \text{rank}(A^t) = M$. Also, since P is an orthogonal projection, each diagonal entry $P_{j,j}$ satisfies $0 \leq P_{j,j} \leq 1$. Furthermore, since no column of A is zero, we actually have $0 < P_{j,j}$ for each j.

Indeed, since $\text{rank}(A) = M$, AA^t is positive definite, and $P = A^t(AA^t)^{-1}A$. Therefore,

$$P_{j,j} = \vec{e}_j \cdot A^t(AA^t)^{-1}A\vec{e}_j = \vec{a}_j \cdot (AA^t)^{-1}\vec{a}_j > 0 \ .$$

Hence, if we define p_j° by

$$\frac{1}{p_j^\circ} = P_{j,j} = \vec{a}_j \cdot (AA^t)^{-1}\vec{a}_j > 0 \ ,$$

we have that whenever $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ is properly redundant, $1 < p_j^\circ < \infty$ for each j, and also $\sum_{j=1}^N (1/p_j) = Tr(P) = M$, so that (4.1) is satisfied. Morover, the Euler–Lagrange equation (3.12) is then satisfied with $S = I$.

Definition Let $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$ be a properly redundant spanning set of vectors in $\mathbb{R}^M$, and let $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N]$. Let P be the orthogonal projection in $\mathbb{R}^N$ onto the image of A^t. For $j = 1, 2, \ldots, N$, define $p_j^\circ = 1/P_{j,j}$. Then $\vec{p}^{\,\circ} = \{p_1^\circ, p_2^\circ, \ldots, p_N^\circ\}$ is the *canonical set of L^p indices corresponding to* $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$. The terminology will be justified by Theorem 6.1 below.

Since for $\vec{p} = \vec{p}^{\,\circ}$, the Euler–Lagrange equations (3.12) are satisfied with $S = I$, it follows from (3.19) that

$$D(\vec{p}^{\,\circ}) = \prod_{j=1}^N (P_{j,j})^{-P_{j,j}/2} \det(AA^t)^{-1/2} \ . \tag{6.1}$$

Notice that while computing $D(\vec{p})$ for given L^p indices is a nonlinear optimization problem, calculating the $D(\vec{p}_0)$ is a simple matter of linear algebra. This is significant since it turns out that given the vectors $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_N\}$, $D(\vec{p}^{\,\circ})$ is the "best best constant" in the generalized Young's inequality

$$\int_{\mathbb{R}^M} \prod_{j=1}^N f_j(\vec{a} \cdot x) \mathrm{d}^N x \leq D \prod_{j=1}^N \|f_j\|_{p_j}$$

This justifes the terminology "canonical L^p indices":

Theorem 6.1: *For any properly redundant spanning set, $D(\vec{p}^{\,\circ}) < D(\vec{p})$ for all $\vec{p} \neq \vec{p}^{\,\circ}$.*

Proof: Let ϕ_A be the function defined by (3.9), and ϕ_A^* its Legendre transform. It was shown by Brascamp and Lieb that ϕ_A is convex. Since ϕ_A is smooth as well as convex, ϕ_A^* is strictly convex. From (3.10) and (4.10), we have

$$2\ln(\tilde{D}(\vec{p})) = \phi_A^*\left(\frac{1}{p_1}, \frac{1}{p_2}, \ldots, \frac{1}{p_N}\right) \tag{6.2}$$

By the definition of ϕ_A, and the Euler–Lagrange equation (3.12), if $1/p_j^\circ$ is the jth canonical L^p index,

$$\left(\frac{1}{p_1^\circ}, \frac{1}{p_2^\circ}, \ldots, \frac{1}{p_N^\circ}\right) = \nabla\phi_A(0) \ . \tag{6.3}$$

But since the gradients of Legendre transforms are inverse to one another,

$$\nabla\phi_A^*(\nabla\phi_A(0)) = 0 \ .$$

This proves that the vector on the left in (6.3) is a critical point of ϕ_A^*. Since ϕ_A^* is strictly convex, it is the unique minimizer. ■

We also note that formula (6.2) displays $D(\vec{p})$ as a log–convex function of $\vec{p}$, This can be used to produce *arbitrarily sharp* upper bounds on $D(\vec{p})$ for a given set of L^p indices: Using Newton's method or some other means of generating explicit approximate solutions of the Euler–Lagrange equations (3.12), generate several approximate solutions. For each, compute the "best best constant" for each $\{s_1\vec{a}_1, s_2\vec{a}_2, \ldots, s_N\vec{a}_N\}$. If $\vec{p}$ can be written as a convex combination of the corresponding vectors of cannionical inverse L^p indices, then $D(\vec{p})$ can be bounded above by a convex combination of the corresponding "best best constants".

Special cases of the canonical L^p indices have arisen in applications of the Brascamp Lieb inequality. A beautiful application to convex geometry by Keith Ball [1] concerened a situation in which N unit vectors $\vec{u}_1, \ldots, \vec{u}_N$ satisify

$$\sum_{j=1}^{N} c_j \vec{u}_j \vec{u}_j^t = I_{M\times M} \ . \tag{6.4}$$

where the c_j are positive numbers. Clearly, $\sum_{j=1}^N c_j = M$. Let $\vec{a}_j = \sqrt{c_j}\vec{u}_j$. Then (6.4) becomes $AA^t = I_{M\times M}$. It follows that the orthogonal projection onto the image of A^t is simply A^tA, and the jth diagonal entry is c_j. Hence taking $p_j = 1/c_j$ gives the canonical L^p indices in this case. These were the L^p indices used by Ball in his application.

Since for the canonical L^p indices, the Euler–Lagrange equation (3.12) is then satisfied with $S = I$, the heat flow interpolation argument of Section 3 gives an especially simple

proof of the inequality in this case. For this reason, the method of proof developed here works very simply in Keith Ball's context; see [4] for more information.

Example: Consider the five vectors

$$\vec{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad \vec{a}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad \vec{a}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \qquad \vec{a}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \vec{a}_5 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \ .$$

It is easily seen that this is a properly redundant spanning set. Notice that the first three vectors all lie in the plane $x_1 + x_2 + x_3 = 0$. As long as $0 < 1/p_j < 1$ for each j,

$$\sum_{j \in S} \frac{1}{p_j} < |S| \ . \tag{6.5}$$

Therefore, as long as $r(S) \geq \min\{|S|, M\}$ and (4.1) is satisfied, and there are no supercritical sets. The only set S with $r(S) < \min\{|S|, M\}$ is $S = \{1, 2, 3\}$. Therefore, as long as $\dfrac{1}{p_1} + \dfrac{1}{p_1} + \dfrac{1}{p_1} < 2$, together with the scaling condition (4.1) and $0 < p_j < 1$ for each j and are all satisfied, $\vec{p}$ belongs to K_A°, and K_A is the closure of the points obtained in this way. An easy computation shows that the canonical indices for this example are $p_1 = 2$, and $p_2 = p_3 = p_4 = p_5 = 8/5$. By Theorem 4.4, K_A is has 9 vertices, and is their convex hull.

Appendix

In this section we exhibit trial functions that show the optimality of Theorem 1 and Theorem 2, and describe the change of variables leading to (5.1).

First we show that the inequality in Theorem 1 cannot hold with *any* constant if the index p of the L^p norms on the right side is less than 2. For any given $0 < \alpha < 1$, define $f(v)$ be defined by

$$f(v) = |v|^{-\alpha} + (1 - v^2)^{-\alpha(N-1)/2} \ . \tag{7.1}$$

Then $\displaystyle\int_{[-1,1]} f^p \mathrm{d}\nu_N < \infty$ as long as $p\alpha < 1$, as one easily sees from (1.10).

On the other hand, discarding one term in each factor,

$$\prod_{j=1}^{N} f(v_j) \geq \left(\prod_{j=1}^{N-1} |v_j|^{-\alpha} \right) (1 - v_N^2)^{-\alpha(N-1)/2} \ .$$

We can parameterize the upper and lower hemispheres of S^{N_1} using the coordinates $(v_1, \ldots, v_{N-1})$ The intergal over S^{N-1} is then easily converted into an integral over the

unit ball in $\mathbb{R}^{N-1}$. Doing this in radial coordinates, we have, since $|v_n| = \sqrt{1-r^2}$ in these coordinates,

$$\int_{S^{N-1}} \left(\prod_{j=1}^{N} f(v_j) \right) \mathrm{d}\mu \geq C \int_0^1 r^{-2\alpha(N-1)} \frac{r^{N-2}}{\sqrt{1-r^2}} \mathrm{d}r$$

where C is a positive constant resulting from the angular integration. This integral diverges unless $\alpha < 1/2$.

The conclusion is that for all N and all $p < 2$, there is a positive function f so that $\int_{S^{N-1}} f^p(v_1) \mathrm{d}\mu < \infty$ while $\int_{S^{N-1}} \left(\prod_{j=1}^{N} f(v_j) \right) \mathrm{d}\mu = \infty$.

Next we turn to the entropy inequality in Theorem 2. Consider a spherical cap on the sphere S^{N-1} centered at the point $v_1 = 1, v_2 = 0, \ldots, v_N = 0$ of radius ε denote its characteristic function by χ_ε. Define

$$F = H\chi_\varepsilon \qquad \text{with} \qquad H = \left(\int_{S^{N-1}} \chi_\varepsilon \mathrm{d}\mu \right)^{-1} .$$

Clearly, H is of order $\varepsilon^{-(N-1)}$, and hence for ε small, $S(F)$ is of order

$$-\log(H) \tag{7.2}$$

which is of order $(N-1)\log(\varepsilon)$. Since the function is invariant under all rotation that fix the v_1 axis, we get that the entropy of the marginal is also given by (7.2). Moreover, the j-th marginal can be thought of as averaging the function $H\chi_\varepsilon$ over all roations that keep the axis v_j fixed. the resulting function is essentially a multiple of a characteristic function of a band of width 2ε that is centered at the equator perpendicular to the v_j axis. Call this function $f_j := \psi_\varepsilon$. Since the integral of this function must be equals to one the height of this function must be $h = (\int_{S^{N-1}} \psi_\varepsilon \mathrm{d}\mu)^{-1}$, and is of order $1/\varepsilon$. Hence its entropy is of order $\log(\varepsilon)$. Thus the sum of the entropies of the marginals is given, in leading order, by $2(N-1)\log(\varepsilon)$ which is twice the entropy of the function F. This shows that the constant 2 in the entropy inequality is sharp.

Finally, the coordinate change leading to (5.1) may be described as follows: Suppose that $\vec{a}_1$ is not in the span of $\{\vec{a}_2, \ldots, \vec{a}_N\}$. Let $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_M\}$ be an orthonormal basis of $\mathbb{R}^M$ so that $\{\vec{u}_2, \ldots, \vec{u}_M\}$ has the same span as $\{\vec{a}_2, \ldots, \vec{a}_N\}$. Let R be the matrix given by $R = [\vec{a}_1, \vec{u}_2, \ldots, \vec{u}_M]$. (That is, the first column of R is $\vec{a}_1$, the second column is $\vec{u}_2$, and so forth). Then R is invertible, and we can define new coordinates z by $z = R^t x$. With this definition, $z_1 = \vec{a}_1 \cdot x$. Moreover, for $j \geq 2$,

$$\vec{a}_j \cdot x = \vec{a}_j \cdot (R^t)^{-1} z = (R^{-1}\vec{a}_j) \cdot z \ .$$

Since $R^{-1}\vec{a}_j$ is the coordinate vector of $\vec{a}_j$ with respect to the basis $\{\vec{a}_1, \vec{u}_2, \ldots, \vec{u}_M\}$, $(R^{-1}\vec{a}_j)_1 = 0$ for $j \geq 2$. Therefore, defining $\vec{w}$ in $\mathbb{R}^{M-1}$ by $w_j = z_{j+1}$, there are uniquely

determined vectors $\vec{b}_j$ in $\mathbb{R}^{M-1}$ so that $(R^{-1}\vec{a}_j)\cdot z = \vec{b}_j \cdot \vec{w}$. Since

$$\mathrm{d}^M x = \frac{1}{|\vec{u}_1 \cdot \vec{a}_1|}\mathrm{d}^M z = \frac{1}{|\vec{u}_1 \cdot \vec{a}_1|}\mathrm{d}z_1 \mathrm{d}^{M-1} w \ ,$$

we have (5.1).

Bibliography

[1] K. Ball: "Volume Ratios and a reverse isoperimetric inequality" Jour. London Math. Soc., **44** no. 2 351–359 (1991)

[2] F. Barthe: "Optimal Young's inequality and its converse, a simple proof" Geom. Func. Analysis., **80** 234–242 (1998)

[3] F. Barthe: "On a reverse form of the Brascamp–Lieb inequality" Invent. Math., **134** no. 2, 235–361 (1998)

[4] F. Barthe and D. Cordero–Erausquin: "Inverse Brascamp–Lieb inequalities along the heat equation" to appear in *Geometric Aspects of Functional Analysis, 2002–2003*, eds. V. Milman and G. Schechtman, Lecture Notes in Mathematics 1850, Springer Verlag, Berlin 2004

[5] H. J. Brascamp and E. H. Lieb: "Best constants in Young's inequality, its converse, and its generalization to more than three functions" Advances in Math., **20** 151-173 (1976)

[6] E. A. Carlen, M. C. Carvalho and M. Loss: Many body aspects of approach to equilibrium, in Journes Equations aux derivees partielles, Nantes, 5-9 juin 2000.

[7] E. A. Carlen, M. C. Carvalho, M. Loss, "Determination of the spectral gap for Kac's master equation and related stochastic evolution", Acta Mathematica **191**, 1-54 (2003). arXiv math-ph/0109003

[8] E. H. Lieb, "Gaussian kernels have only Gaussian maximizers", Invent Math **102**, 179–208 (1990).